Attorney Docket No.: 5442.400-US PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Outtrup et al.

Serial No.: 09/382,096 Group Art Unit: 1652

Filed: August 24, 1999 Examiner: Patterson, C.

For: Alkaline Bacillus Amylase

DECLARATION OF PREBEN NIELSEN

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

I, Preben Nielsen , do hereby state and declare that

1. I am a citizen of Denmark residing at Hoersholm, Denmark.

2. I am a Science Manager at Novozymes A/S. I have a Masters of Science degree in Biochemistry from Copenhagen University (Copenhagen, Denmark) and a PhD in Biotechnology from Heriot Watt University (Edinburgh, UK). I have expertise in microbiology and systematic biotechnology, as well as in production, detection and characterization of extra-cellular enzymes. Since 1995, I have worked with screening and production of bacterial enzymes.

3. I have read and am familiar with Boyer et al. (US Pat. No. 4,061,541). I was requested to make a comparison between the alpha-amylase obtained from the bacterium NRRL 3881, which is described in Boyer et al., with the alpha-amylase produced by Bacillus sp. NCIMB 40916, having an amino acid sequence as shown in positions 1-556 of SEQ ID NO:4 of the above-captioned U.S. patent application.

As discussed below, it is my conclusion that the alpha-amylase stated by Boyer et al. to be produced by the bacterium NRRL 3881 would not be identical to the alpha-amylase produced by the bacterium Bacillus sp. NCIMB 40916 because, among other things, the bacterium NRRL 3881 is phylogenetically different from the bacterium Bacillus sp. NCIMB 40916. Moreover, despite following highly specific, art recognized procedures, we were unable to show the presence of a detectable amount of alpha-amylase activity from NRRL 3881, and it is my conclusion that NRRL 3881 does not produce an alpha-amylase

4. We have performed a classification of the bacterium NRRL 3881, and based on the 16S rRNA gene, we concluded that NRRL 3881 is a strain of Bacillus halodurans.

5. We obtained a sample of the bacterium NRRL 3881 and attempted to produce the alpha-amylase alleged to be produced from NRRL 3881 by U.S. Patent No. 4,061,541 in a method equivalent to the method described in U.S. Patent No. 4,061,541 for comparison to the alpha-amylase from Bacillus sp. NCIMB 40916. Specifically, we did a Phadebas assay (Pharmacia), which is standard assay known in the art for analyzing amylolytic activity. This assay is particularly good for alpha-amylase activity, but side activity will be seen from other amylolytic enzymes, such as CGTases and alpha-glycosidases. This assay is a very sensitive assay, and would detect the alpha-amylase activity if it was present.

We fermented the NRRL 3881 strain and ran the Phadebas assay twice (pH 9.5), and both times we were unable to produce sufficient amylase to make a proper pH profile. Based on this test, we concluded that NRRL 3881 does not produce an alpha-amylase. This conclusion is consistent with classification information, which shows that NRRL 3881 is a Bacillus halodurans strain. Bacillus halodurans is reported in the literature to produce straight starch degrading enzyme activity, but from genome sequence it is known not to produce alpha-amylase. See Preben Nielsen et al., "Phenetic diversity of alkaliphilic Bacillus strains: proposal for nine new species," Microbiology 141, 1745-1761 (1995) and Takami et al. "Complete Genome Sequence of the Alkaliphilic Bacterium Bacillus halodurans and Genomic Sequence Comparison with Bacillus subtilis," Nucleic Acid Research, 28, 4317-4331 (2000).

This conclusion is also supported by the PCR reaction we performed on NRRL 3881. In particular, we made a PCR reaction using two degenerated primers that recognize the gene of known alpha-amylases from bacillus, and were unable to detect a product with DNA from NRRL 3881.

6. The undersigned declarant declares further that all statements made herein of his own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize any patent issuing thereon.

Signed this day 25 of August, 2003

Preben Nielsen

Microbiology (1995), 141, 1745–1761

Printed in Great Britain

Phenetic diversity of alkaliphilic *Bacillus* strains: proposal for nine new species

Preben Nielsen,[1,2]† Dagmar Fritze[1] and Fergus G. Priest[2]

Author for correspondence: Fergus G. Priest. Tel: +44 131 449 5111. Fax: +44 131 451 3009.

[1] DSM – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany

[2] Department of Biological Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

One hundred and nineteen strains of alkaliphilic and alkalitolerant, aerobic endospore-forming bacteria were examined for 47 physiological and biochemical characters, and DNA base composition. Numerical analysis (S_j and S_{sm}/UPGMA clustering) revealed 11 clusters that comprised three or more strains. Most of the phena were further characterized by analysis of carbohydrate utilization profiles using the API 50CH system, but strains of two taxa could not be cultured by this method. DNA reassociation studies showed that nine of the phena were homogeneous, but strains of phenon 4 and phenon 8 were each subdivided into two DNA hydridization groups. The strains could therefore be classified into 13 taxa plus a number of unassigned single-membered clusters. Two taxa were equated with *Bacillus cohnii* and *B. alcalophilus* and nine of the remainder are proposed as new species with the following names: *B. agaradhaerens* sp. nov., *B. clarkii* sp. nov., *B. clausii* sp. nov., *B. gibsonii* sp. nov., *B. halmapalus* sp. nov., *B. halodurans* comb. nov., *B. horikoshii* sp. nov., *B. pseudalcalophilus* sp. nov. and *B. pseudofirmus* sp. nov. Two taxa were insufficiently distinct to allow confident identification and these have therefore not been proposed as new species.

Keywords: alkaliphile, *Bacillus*, classification, nomenclature, taxonomy

INTRODUCTION

During an attempt to improve the enrichment medium for *Vibrio cholerae*, Vedder (1934) isolated aerobic, endospore-forming bacteria from human faeces, and later from animal faeces, which proved to be obligately alkaliphilic organisms (defined as a pH optimum for growth above pH 9 and no growth at pH 7). He proposed the name *Bacillus alcalophilus* for his strains and stated that he had been able to prove that life exists which not only tolerates, but depends on, a highly alkaline pH. Today, these and other alkaliphilic *Bacillus* strains are of considerable industrial interest, particularly for the production of enzymes such as proteases for inclusion in laundry detergents (Aunstrup *et al.*, 1972), xylanases for use in the pulp paper industry (Nakamura *et al.*, 1993) and cyclodextrin glucanotransferase for cyclodextrin manufacture from starch (Kitamoto *et al.*, 1992). These industrial applications have prompted the isolation of strains from a variety of alkaline environments (Horikoshi, 1991; Jones *et al.*, 1994). The diverse bacteria recovered are described simply as *Bacillus* sp. because there is no taxonomic framework to enable identification.

In addition to strict alkaliphiles, some *Bacillus* species are reported to tolerate a more or less alkaline pH and may be described as alkalitolerant (defined as growth at or above pH 9 but also at pH 7). The asporogenous bacterium *B. halodenitrificans* (Denariaz *et al.*, 1989) is reported to grow between pH 5·8 and 9·6, but has an optimum in the neutral range at pH 7·4, and two thermophilic *Bacillus* species are reported to actually require a slightly alkaline pH: '*B. pallidus*' (Scholz *et al.*, 1988), which grows best at pH 8–8·5 while also tolerating pH 7; and '*B. thermocloacae*' (Demharter & Hensel, 1989), which requires a pH of 8–9 and does not grow at pH 7.

Taxonomic studies of truly alkaliphilic *Bacillus* strains have been few. Boyer *et al.* (1973) deemed two halotolerant and alkaliphilic *Bacillus* strains to be sufficiently related to the type strain of *B. alcalophilus* to justify subspecies status as '*B. alcalophilus* subsp. *halodurans*', which necessitated transforming *B. alcalophilus* to '*B. alcalophilus* subsp. *alcalophilus*'. This proposal was not included in the Approved Lists of Bacterial Names (Skerman *et al.*, 1980). In later studies, strains of '*B. alcalophilus* subsp. *halodurans*' were shown to be so different from the parent species, both genetically and physiologically, that separate species status

†**Present address:** Microbiology Section, The Royal Veterinary and Agricultural University, DK-1958 Fredericksberg, Copenhagen, Denmark.

Abbreviation: MUG, 4-methylumbelliferyl β-D-glucuronide.

was considered more appropriate (Fritze *et al.*, 1990). Boyer *et al.* (1973) considered their strains to be distantly related to an alkaliphilic *B. circulans* strain described by Chislett & Kushner (1961). This organism (strain RU 38) was later classified in a newly established alkaliphilic taxon, *Bacillus cohnii* (Spanka & Fritze, 1993).

Some indication of the diversity within alkaliphilic *Bacillus* strains was provided by Gordon & Hyde (1982). They characterized 174 alkaliphilic *Bacillus* strains of industrial importance after adapration of the strains to neutral pH and allocated them to five physiologically defined groups. Four of these groups were sufficiently homogeneous to be retrieved in a later investigation with only a few compositional changes (Fritze *et al.*, 1990). More recently, a strain of an alkaliphilic spore-forming organism (growth at pH 8–10 but not at pH 7) was described which was so peculiar in its properties that the authors established a new genus for it: *Amphibacillus xylanus* (Niimura *et al.*, 1990). Its lack of cytochromes, quinones or catalase and its ability to form spores under aerobic as well as anaerobic conditions clearly distinguished this organism from the genera *Bacillus*, *Clostridium* and *Sporolactobacillus*.

The aim of the present work was to classify, at the species level, a collection of industrially important alkaliphilic *Bacillus* strains. Previous studies (Nielsen *et al.*, 1994) had indicated that some of these bacteria were phylogenetically distinct from all validly described *Bacillus* species with the exceptions of the two alkaliphilic species *B. alcalophilus* and *B. cohnii*.

METHODS

Bacterial strains and culture conditions. The bacterial strains used in this study are listed in Table 1. The strains were stored as a working collection at 5 °C on Nutrient agar (Difco) slopes adjusted to pH 9 with 0·1 M $NaHCO_3$ or to pH 10 with 0·1 M $NaHCO_3$ and 0·1 M Na_2CO_3 (sodium sesquicarbonate) and supplemented with 0·1 % $MnSO_4$ and 0·1 % $MgCl_2$.

Phenotypic tests. Phenotypic tests were performed using the methods of Gordon *et al.* (1973), with media adjusted to approximately pH 10 by the addition of 100 ml l^{-1} of 1 M sodium sesquicarbonate buffer, except for the phenylalanine test, where 50 ml l^{-1} was used (Fritze *et al.*, 1990). Tests were repeated where unclear or unexpected results were obtained.

pH range for growth. Nutrient agar (1 litre) was adjusted to various pH values: pH 6·0 (adjusted by adding HCl), pH 7·0 (addition of 100 ml 1·0 M sodium phosphate buffer, pH 7·0), pH 8·0 (addition of 100 ml 1·0 M Na_2HPO_4), pH 9·0 (addition of 100 ml 1 M $NaHCO_3$), pH 10·0 (addition of 100 ml 1 M sodium sesquicarbonate buffer, pH 10). The agar media and buffers were autoclaved separately. Agar plates were inoculated by streaking to give discrete colonies, and growth intensity was interpreted on a scale from 1–5 after 1, 3 and 5 d. Coding for the numerical analysis is described in the footnote to Table 2.

Utilization of carbohydrates. An early stationary phase culture in Nutrient broth (Difco) was centrifuged, and the cells resuspended in an equal amount of resuspension buffer (2·19 g $Na_2HPO_4.2H_2O$, 0·75 g KH_2PO_4, 1 g NaCl, 0·05 g $MgSO_4.7H_2O$ in 250 ml distilled water). Resuspended culture (0·5 ml) was added to a soft agar held at 42 °C and composed of 4 ml each of two solutions. Solution 1 was KNO_3 (2 g), $Na_2HPO_4.2H_2O$ (2 g), NaCl (5 g), 100 ml 1 M sodium sesquicarbonate buffer (pH 10·0) and water to 500 ml. After autoclaving, 2 ml trace element solution [0·1 g $ZnSO_4.7H_2O$, 0·03 g $MnCl_2.4H_2O$, 0·3 g H_3BO_3, 0·2 g $CoCl_2.6H_2O$, 0·01 g $CuCl_2.2H_2O$, 0·02 g $NiCl_2.6H_2O$ and 0·03 g $Na_2MoO_4.2H_2O$ dissolved in 1 l distilled water with 12·8 g titriplex I (Merck) and 0·1 g $FeCl_3.4H_2O$ added], 2 ml vitamin solution (0·2 mg biotin, 2·0 mg nicotinic acid, 1·0 mg thiamin, 1·0 mg 4-aminobenzoate, 0·5 mg pantothenate, 5·0 mg pyridoxamine, 2·0 mg cyanocobalamin, dissolved in 100 ml water) and 2·5 ml 2 % $MgSO_4.7H_2O$ were added. Solution 2 was agar (3 g) autoclaved in 500 ml distilled water. The API 50CH gallery (Biomérieux) was inoculated with two drops of the culture using a Pasteur pipette. Growth on the various carbohydrates was examined after 2, 7 and 14 d as recommended by the manufacturer.

Numerical analysis. The phenotypic data, except carbohydrate utilization results, were coded for numerical analysis as binary characters. Carbohydrate utilization data were excluded because some strains failed to grow in the test system, thus providing an incomplete data matrix. The pH range for growth was coded as described in the footnote to Table 2 and mol % G+C was coded by additive coding into six characters: > 34, > 36, > 38, > 40, > 42 and > 44 mol % G+C. Data were analysed using the NTSYS software package (Exeter Software). Similarity matrices were calculated using the Simple Matching (S_{SM}) and Jaccard (S_J) coefficients, and UPGMA dendrograms were drawn using the programs SIMEQUAL, and DRAWTREE of NTSYS.

Mol % G+C determination. Minipreparations of DNA were prepared from cell mass (2–3 inoculation loops) taken from agar plates. Cells were washed in 1 ml TES buffer (50 mM Tris, 5 mM EDTA, 2·5 % sucrose, pH 8) and resuspended in 100 µl TES buffer. A further 390 µl TES containing 20 mg lysozyme ml^{-1} was added and the tubes incubated at 37 °C for at least 30 min. Lysis buffer (300 µl 2 mM EDTA, pH 8, containing 60 % guanidinium thiocyanate and 12·5 % SDS) was added followed by 250 µl 7·5 M ammonium acetate. After mixing and chilling on ice for 10 min, the tubes were centrifuged for 15 min and the supernatant removed to a new 2 ml microfuge tube. The preparation was extracted with chloroform/isoamyl alcohol, precipitated with 2-propanol, washed with 70 % ethanol, dried and resuspended in RNase buffer (Seldin & Dubnau, 1985). The DNA was extracted with phenol and then with chloroform/isoamyl alcohol, precipitated in ethanol, washed in 70 % ethanol, dried and resuspended in water. An isolation typically yielded approximately 20 µg DNA and about 90 % of the RNA was removed. Hydrolysis and dephosphorylation of the DNA was carried out as described by Mesbah *et al.* (1989). Nucleoside samples, 25 µl of 100 µg ml^{-1}, were separated by reverse-phase HPLC at 26 °C using 0·6 M $NH_4H_2PO_4$/acetonitrile 80:6 (v/v), pH 4·4, as solvent and a flow rate of 0·7 ml min^{-1} (Tamaoka & Komagata, 1984). The HPLC apparatus consisted of a high-pressure pump and UV detector (LKB 2150 and 2151). Separation was performed through a precolumn and an analytical column (Nucleosil 100-5C18 20 × 4 mm and 250 × 4 mm respectively; MELZ VDS, Berlin). UV absorption was detected and analysed by an integrator (CR-3A, Shimadzu). Mol % G+C was calculated from deoxyguanosine and deoxythymidine contents (Mesbah *et al.*, 1989).

DNA hybridization. DNA for hybridizations was prepared from 300 ml cultures in Tryptone soy broth (Difco) adjusted to pH 9 or 10 as described above using the method of Seldin & Dubnau (1985). The DNA concentration was determined by the diphenylamine assay (Johnson, 1981) with salmon testis DNA in the range 10–100 µg ml^{-1} as standard. Probe DNA was prepared by random-primed labelling of total chromosomal DNA with 5 µl [^{35}S]dATP [600 Ci (22·2 TBq) $mmol^{-1}$, 50 µCi (1·85 MBq)] using the Nonaprimer Kit (Appligene). Target DNA (100 ng per slot) was immobilized on nylon membranes

Table 1. Bacterial strains included in this study

Strain designation					Source§	References‖
PN	Cluster†	NCIMB	DSM	Other		
1	4a	10291	8714	C 339	River bank soil, Denmark	*a, d, e, j*
2	4a	10282		C 323	Field soil, Denmark	*a, d, e*
3	1	10283	8715	C 324	Lake bank soil, Holstein, Germany	*a, d, c, j*
4	7	10284		C 325	Infection on perborate plate	*a, d, e, k*
5	4a	10289		C 337	Clay from field, Holstein, Germany	*a, d, e*
6	1	10290		C 338	Cementary soil, Denmark	*a, d, c*
7	1	10292		C 340	River bank soil, Denmark	*a, d, c*
8	4b	10293		C 341	Field soil, Denmark	*a, d, c*
9	1	10294		C 342	Garden soil, Denmark	*a, d, c*
10	1	10295		C 343	Garden soil, Denmark	*a, d, c*
11	1	10296		C 346	Chicken yard soil, Denmark	*a, d, e*
12	1	10285		C 326	Infection on perborate plate	*a, d, e*
13	1	10297		C 347	Deer manure, Denmark	*a, d, e*
14	1	10298		C 348	Chicken run soil	*a, d, e*
15	1	10299		C 349	Deer manure, Denmark	*a, d, e*
16	1	10300		C 350	Fresh lake water, Denmark	*a, d, e*
17	5*	10301		C 351	Chicken manure	*a, d, e*
18	1	10302		C 352	Ostrich manure, zoo	*a, d, e*
19	1	10303		C 353	Soil and leaves, Denmark	*a, d, e*
20	5	10304		C 354	Chicken yard soil	*a, d, e*
21	1*	10305		C 355	Chicken yard soil	*a, d, e*
22	5*	10306		C 356	Garden bark	*a, d, e*
23	6	10309	8716	C 360	Garden soil, Denmark	*a, d, e, j*
24	6	10317		C 372	Clay from field, Holstein, Germany	*a, d, e*
25	4b*	10281		C 311	Wood soil, Holstein, Germany	*a, d, e*
26	4b	10288	8717	C 336	Horse and elephant manure	*a, d, e, j*
27	1			RAB		*f, d, e*
28	5	10307		C 357	Chicken yard soil	*a, d, e*
29	5	10308		C 358	Chicken yard soil	*a, d, e*
30	5	10310	8718	C 364	Lavatory cistern	*a, d, e, j*
31	5	10311		C 365	Liquid from tannery liming bath	*a, d, e*
32	5	10312		C 366	Baby faeces	*a, d, e*
33	4b	10314		C 369	Ostrich manure, zoo	*a, d, e*
34	7	10318		C 373	Garden soil, Denmark	*a, d, e, k*
35	6	10319		C 374	Clay from field, Holstein, Germany	*a, d, e*
36	5	10320		C 375	Ostrich manure	*a, d, e*
37	5	10321		C 376	Elephant manure	*a, d, e*
38	5	10322		C 377	Water from hippopotamus basin	*a, d, e*
39	5	10323		C 378	Tannery liming bath containers	*a, d, e*
40	5	10324		C 410	Tiger manure	*a, d, e*
41	5	10325		C 411	Pigeon manure	*a, d, e*
42	5	10326		C 412	Chicken yard soil	*a, d, e*
43	6		2512	ATCC 21522, 221		*d, e, g*
44	5		2513	ATCC 21591, A-59		*b, d, e*
45	6*		2514	ATCC 21536, O-4		*d, e, b*
46	6		2515	ATCC 21537, Y-76		*d, e, b*
47	1		2516	ATCC 21592, A-40		*d, e*
48	1		2517	ATCC 21593, 124-1		*d, e*
49	UC		2518	ATCC 21594, 169		*d, e, b*
50	UC		2519	ATCC 21595, 135		*d, e, b*
51	1*		2520	ATCC 21596, 27-1		*d, e*
52	UC		2521	ATCC 21832, N1		*d*
53	UC		2522	ATCC 21833, N4		*d*
54	4b		2523	ATCC 31006, 13		*d, e, b*

[redacted]

Table 1. (cont.)

Strain designation					Source§	References‖
PN	Cluster†	NCIMB	DSM	Other		
55	4b		2524	ATCC 31007, 17-1		*d, e, b*
56	6		2525	ATCC 31084, M-29		*e*
57	10‡	10438	2526	NRS1554, NCTC 4554		*d, e, l*
58	7	9218	2528	RU 38	Contamination, alkaline agar plate	*e, d, e, k*
59	1*	10286	6930	C 334	River bank soil, Denmark	*a, d, e*
60	7	10327	6932	C 413	Clay from field, Holstein, Germany	*a, d, e*
61	5	10313	6939	C 367	Elephant manure	*a, d, e*
62	5	10316	6940	C 371	Elephant manure	*a, d, e*
63	5*		6941	BC 4	K. Aunstrup	*d, e*
64	UC		6942	BC 7	K. Aunstrup	*d, e*
65	UC		6943	PB 9	K. Aunstrup	*d, e*
66	5		6944	BB 16	K. Aunstrup	*d, e*
67	7			5A	Water/sand, Wadi Natrun, Egypt	*k*
68	7			5' A	Water/sand, Wadi Natrun, Egypt	*k*
69	7			14 B	Water/sand, Wadi Natrun, Egypt	*k*
70	7			15 B	Water/sand, Wadi Natrun, Egypt	*k*
71	7			15' B	Water/sand, Wadi Natrun, Egypt	*k*
72	1			AF 1	Stalactite cave, France	
73	1*			AF 2	Stalactite cave, France	
74	UC			Br-B	Soil, Brazil	
75	6*			Br-E	Soil, Brazil	
76	6			Br-F	Soil, Brazil	
77	1		6950	BC 3	K. Aunstrup	*d, e*
78	7		6913	PB 38	K. Aunstrup	*d, e, k*
79	8b		6951	PB 19	K. Aunstrup	*d, e*
80	5		497		*'B. alcalophilus* subsp. *halodurans'*	*b, d, e*
81	1*		7316	O 2	Owens Lake, California, USA	*d*
82	10		7317	O 3	Owens Lake, California, USA	*d*
83	7		7318	M 5	Mono Lake, California, USA	*d*
84	UC		7319	M 8	Mono Lake, California, USA	*d, k*
100	10*			JA 14	H. Outtrup	
101	10			JA 16	H. Outtrup	
102	3		8720	JP 395	H. Outtrup	*j*
103	3			AC 66	H. Outtrup	
104	3			K 51	H. Outtrup	
105	2		8721	AC 13	H. Outtrup	*j*
106	2			K 320	H. Outtrup	
107	2			K 316	H. Outtrup	
108	2			AC 59	H. Outtrup	
109	11		8722	PD 138	H. Outtrup	*j*
110	11			H 55	H. Outtrup	
111	11			DP 182	H. Outtrup	
112	11			J 102	H. Outtrup	
113	11*			J 114	H. Outtrup	
114	UC			J 20	H. Outtrup	
117	8b*			SF 16	H. Outtrup	
118	8a		8723	JP 170	H. Outtrup	*j*
119	8b			874 X	H. Outtrup	
120	8b			JP 75	H. Outtrup	
121	8b		8719	JP 277	H. Outtrup	*j*
122	8b			MHas96	H. Outtrup	
123	8b*			PD 456	H. Outtrup	
124	8a			JP 216	H. Outtrup	
125	UC			BC 7	K. Aunstrup	

Table 1. *(cont.)*

Strain designation					Source§	References‖
PN	Cluster†	NCIMB	DSM	Other		
127	2			AT 67	H. Outtrup	
129	2			AP 72	H. Outtrup	
130	9			DP 155	H. Outtrup	
131	10*		8724	DP 45	H. Outtrup	*j*
132	9			DP 100	H. Outtrup	
133	11			J 26	H. Outtrup	
135	9			J 463	H. Outtrup	
136	9*			J 466	H. Outtrup	
137	9		8725	DP 44	H. Outtrup	*j*
138	10			DP 469	H. Outtrup	
139	10			DP 486	H. Outtrup	

† Strains marked by an asterisk were included in the given cluster by S_J/UPGMA analysis but showed less than 65 % DNA hybridization to the reference strain. UC, unclustered.

‡ This strain was recovered as a single-membered cluster by S_J/UPGMA analysis but DNA hybridization revealed that it was a member of this taxon.

§ K. Aunstrup and H. Outtrup, Novo Nordisk A/S, Bagsvaerd, Denmark.

‖ References: *a*, Aunstrup *et al.* (1971); *b*, Boyer *et al.* (1973); *c*, Chislett & Kushner (1961); *d*, Fritze *et al.* (1990); *e*, Gordon & Hyde (1982); *f*, Guffanti (1983); *g*, Horikoshi (1971); *h*, Horikoshi (1975); *i*, Horikoshi & Ikeda (1997); *j*, Nielsen *et al.* (1994); *k*, Spanka & Fritze (1993); *l*, Vedder (1934).

(Nytran N, Schleicher and Schuell) using the slot-blot system. Hybridizations were carried out overnight at 62 °C (Alexander & Priest, 1989). Percentage reassociation was calculated from the radioactivity of the hybrids relative to homologous controls and with subtraction of unspecific background radiation occurring from non-homologous (salmon testis DNA) hybridization (Seldin & Dubnau, 1985).

RESULTS

Phenetic characterization

Our initial studies revealed that the alkaliphilic strains included in this study did not produce acid from glucose when examined using the classical test procedures (Gordon *et al.*, 1973) adapted for use at alkaline pH. Sugar fermentation tests were therefore excluded from the study. Moreover, some strains, notably those of cluster 3 and cluster 7 (*B. cohnii*), failed to grow or produce a reliable indication of carbon utilization in the modified API 50CH tests; since these data were incomplete they were excluded from the numerical analysis. The numerical classification was therefore based on 47 characters including DNA base composition, which was coded as six binary characters as described in Methods. Most of these tests had previously been shown to be useful for classification of alkaliphilic *Bacillus* strains (Fritze *et al.*, 1990).

Despite the relatively few tests, the 119 test strains were consistently allocated to 11 clusters in analyses using the Jaccard (at the 80 % similarity level) and simple matching (at the 90 % similarity level) coefficients with average linkage (UPGMA) cluster analysis. The only differences between the two classifications were that strain PN-10 clustered with phenon 1 and strain PN-82 with phenon 10 in the S_J dendrogram but both were recovered as single-membered clusters by S_{SM} analysis. Subsequent evaluation of these results by DNA hybridization revealed that the S_J allocation was correct for strain PN-10 (DNA could not be prepared from strain PN-82) and therefore this classification is shown in Fig. 1. Nine strains were recovered as single-membered clusters and two strains (PN-64 and PN-125) as a doublet. Biochemical and physiological characteristics of clusters 1–11 are shown in Table 2.

Determination of carbohydrate utilization required development of a minimal medium. Ammonium was ineffective as a nitrogen source, probably because ammonia gas was released at the high pH, and therefore nitrate was used in a minimal medium (see Methods) at pH 10. The patterns of carbon source utilization are shown in Table 3. Utilization of aesculin could not be detected because the substrate was unstable under the alkaline conditions. Eight carbohydrate substrates were catabolized by essentially all strains tested. These were: cellobiose, fructose, glucose, glycerol (with the exception of group 8a), maltose, mannitol, sucrose and trehalose. Conversely, no growth was obtained for any strains on the following substrates: adonitol, arbutin, L-arabitol, D-fucose, L-fucose, inulin, 5-ketogluconate, sorbose and L-xylose. The percentage distribution of positive characters for the taxa for the remaining 31 substrates are listed in Table 3.

The clusters derived by phenotypic analysis were, with the exception of *B. cohnii* (phenon 7), which had been previously studied by Spanka & Fritze (1993), evaluated by DNA hybridization. The same reference strains were

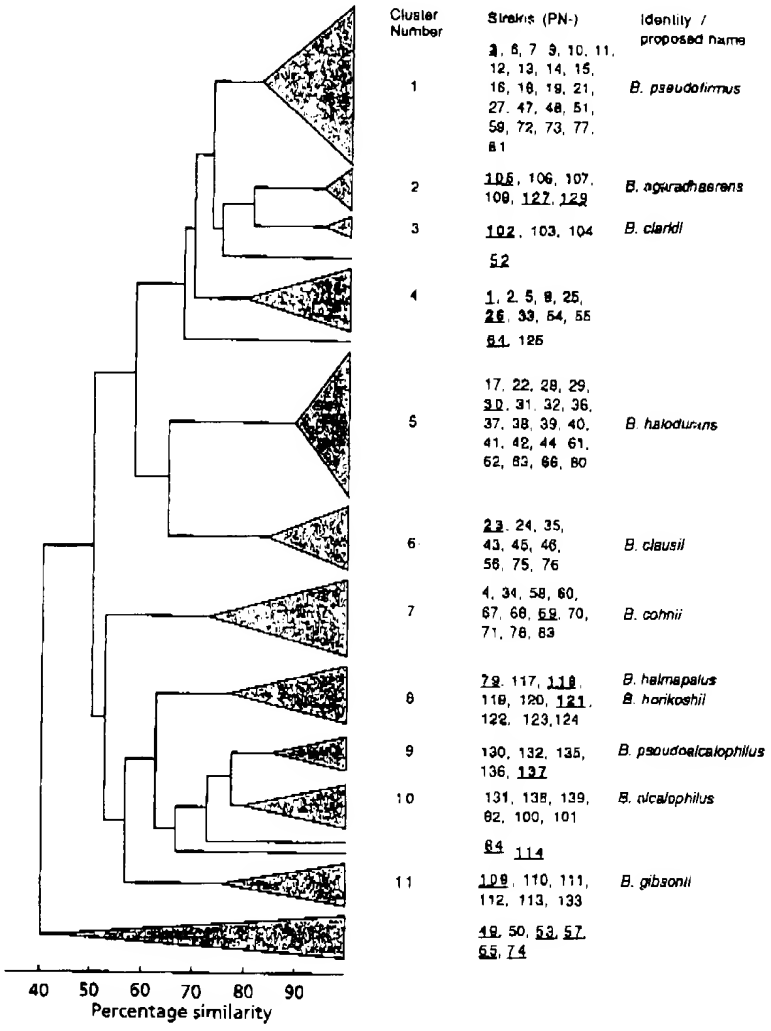


Fig. 1. Abbreviated dendrogram showing the allocation of strains to clusters based on the S_j/UPGMA analysis of the biochemical and physiological characteristics. Bold underlined strain numbers indicate those strains for which the full 16S rRNA sequence has been determined (Nielsen *et al.*, 1994); underlined strain numbers indicate those for which a partial 16S rRNA sequence has been determined (unpublished).

used as in the previous rRNA sequence analysis study of these bacteria (Nielsen *et al.*, 1994). The rRNA study had indicated that clusters 4 and 8 may be heterogeneous and therefore two reference strains were chosen for each of these clusters (see Table 4). In general, the classifications based on phenetic analysis and DNA hybridization were consistent, although there were some discrepancies, as described below and in the descriptions of phena.

DNA from none of the reference strains hybridized significantly with DNA from any other reference strain, thus supporting the integrity of the 11 clusters. Clusters 4 and 8 were each shown to represent at least two DNA hybridization groups, which are consequently labelled a and b in Tables 1, 2 and 3. DNA from only one of the single-membered clusters hybridized significantly with DNA from any of the reference strains (data not shown), supporting the recovery of these strains outside the 11 clusters in the numerical analysis. The exception was strain PN-57, which showed high binding to DNA from *B. alcalophilus* DSM 485^T and was thus considered as a member of cluster 10.

Analysis and descriptions of the groupings of the alkaliphilic *Bacillus* strains

Phenon 1. This large group of 23 strains, derived from

Table 2. Biochemical and physiological characteristics of clusters shown in the S_j/UPGMA analysis modified according to the DNA hybridization results

Cluster ...	1	2	3	4a*	4b	5	6	7	8a*	8b	9	10	11
No. of strains ...	18	4	3	3	5	16	6	11	2	4	4	5	5
Hydrolysis of:													
Hippurate	0	0	100	0	20	6	0	100	100	100	0	0	0
MUG	0	0	0	0	0	0	0	100	0	0	0	0	100
Pullulan	11	50	0	60	100	100	0	100	100	100	100	100	0
Starch	100	100	0	100	100	100	100	100	100	100	100	100	0
Tween 20	0	0	0	0	0	12	0	0	0	0	0	20	0
Tween 40	100	100	100	0	0	100	0	100	0	75	75	100	60
Tween 60	100	100	100	0	0	100	0	100	0	75	75	100	60
Deamination of:													
Phenylalanine	100	0	0	0	0	0	0	0	0	0	0	0	0
Reduction of:													
Nitrate	0	100	100	0	0	12	100	100	0	0	0	20	40
pH optimum†													
7	11	0	0	100	100	86	100	54	100	100	0	0	100
> 9	100	100	100	100	100	100	0	100	0	0	100	100	0
10	17	100	100	67	0	95	0	0	0	0	100	80	0
> 10	6	100	100	0	0	0	0	0	0	0	0	0	0
Growth at:													
10 °C	100	100	0	100	100	0	0	78	100	100	100	100	80
40 °C	100	100	100	0	100	100	100	100	100	100	100	100	0
45 °C	100	100	100	0	0	100	100	100	0	0	0	0	0
50 °C	0	0	0	0	0	100	100	0	0	0	0	0	0
55 °C	0	0	0	0	0	100	0	0	0	0	0	0	0
Growth in NaCl:													
5 %	100	100	100	100	100	100	100	100	0	100	100	100	100
7 %	100	100	100	100	100	100	100	100	0	100	100	100	100
8 %	100	100	100	100	100	100	100	36	0	100	100	100	100
9 %	100	100	100	100	100	100	84	36	0	50	100	0	100
10 %	100	100	100	100	100	100	84	0	0	0	100	0	40
11 %	100	100	100	100	100	100	0	0	0	0	0	0	20
12 %	100	100	100	100	100	100	0	0	0	0	0	0	20
13 %	100	100	100	100	100	35	0	0	0	0	0	0	0
14 %	100	100	100	100	100	0	0	0	0	0	0	0	0
15 %	100	100	100	100	100	0	0	0	0	0	0	0	0
16 %	100	100	100	0	20	0	0	0	0	0	0	0	0
17 %	78	25	67	0	0	0	0	0	0	0	0	0	0
18 %‡	0	–	33	0	0	0	0	0	0	0	0	0	0

All strains were positive for the following tests: casein and gelatin hydrolysis, growth at 15, 20, 30 and 37 °C, and pH optimum for growth > 8 (see below). They were all negative in the following reactions: hydrolysis of Tween 80, and growth at 50 °C and in 20 % NaCl (strains PN-64 and PN-125 positive for the last mentioned).

* Phena 4 and 8 split into two subclusters as indicated by DNA hybridization results. All other clusters (with the exception of cluster 7) are based on strains with > 65 % homology to the reference strain.

† pH 'optimum' for growth coded as positive in '> 8' indicates that growth at pH 8 was estimated to be greater than at pH 7 (all strains displayed this attribute) and '> 9' indicates growth at pH 9 greater than at pH 8. '10' and '> 10' indicate that growth at pH 10 was greater than at pH 8 and pH 9 respectively. For pH 7, a positive record indicates growth at neutrality.

‡ A dash indicates no data available for that entry.

soils from northern Europe, formed a discrete cluster at 82 % S_j. Eighteen of these strains showed > 65 % DNA hybridization with strain PN-3 whereas four strains hybridized to a lesser extent with the reference strain (see Table 4). The characteristics of these bacteria are shown in Tables 2 and 3. They were particularly notable for their

Table 3. Utilization of carbohydrate substrates by strains of clusters shown in the S_j/UPGMA analysis modified according to the results of DNA/DNA hybridization

Substrate	Cluster ...	1	2	4a	4b	5	6	8a	8b*	9	10†	11
	No. of strains ...	18	4	3	5	16	6	2	3	4	3	5
		Percentage positive										
2	Erythritol	0	0	0	0	0	50	0	0	0	0	0
3	D-Arabinose	22	100	0	20	44	50	0	0	25	67	20
4	L-Arabinose	5	100	0	40	94	100	0	0	100	100	100
5	Ribose	89	100	0	40	100	100	0	33	100	100	100
6	D-Xylose	55	100	0	20	100	100	50	33	100	33	40
9	Methyl β-D-xyloside	0	0	0	0	31	33	0	0	75	100	40
10	Galactose	5	100	33	60	100	83	0	0	100	100	40
13	Mannose	55	100	33	60	100	100	100	33	0	33	100
15	Rhamnose	0	60	0	0	100	100	0	0	75	100	40
16	Dulcitol	0	0	0	20	0	100	0	0	0	0	0
17	Inositol	5	0	0	60	100	17	0	0	0	33	0
19	Sorbitol	0	50	0	0	0	100	0	0	0	0	0
20	Methyl α-D-mannoside	0	25	33	0	0	67	0	0	0	33	40
21	Methyl α-D-glucoside	17	100	33	20	94	100	0	67	100	100	0
22	*N*-Acetylglucosamine	100	100	33	80	100	100	100	100	0	100	0
23	Amygdalin	22	100	33	80	100	100	100	33	100	100	60
26	Salicin	67	100	67	60	100	100	50	0	100	100	100
29	Lactose	0	50	0	0	100	83	0	0	100	100	100
30	Melibiose	0	100	33	60	100	100	0	0	0	100	100
34	Melezitose	0	25	33	60	100	100	0	0	0	33	100
35	Raffinose	0	100	33	60	94	100	0	0	0	100	100
36	Starch	95	100	0	20	100	100	100	100	100	100	0
37	Glycogen	67	100	0	0	100	100	100	100	100	100	0
38	Xylitol	5	0	0	20	81	100	0	0	0	0	0
39	Gentiobiose	11	10	0	20	56	83	0	67	100	100	100
40	D-Turanose	39	100	0	20	94	100	100	67	100	100	100
41	D-Lyxose	22	100	0	20	87	100	0	0	75	67	0
42	D-Tagatose	0	75	0	0	6	100	0	0	75	67	0
45	D-Arabitol	16	75	0	40	75	83	0	0	0	33	0
47	Gluconate	61	0	0	20	100	17	50	67	25	0	0
48	2-Ketogluconate	5	100	33	20	100	83	50	67	50	0	20

All strains assigned to clusters were positive for the following substrates: cellobiose, fructose, glucose, glycerol, maltose, sucrose and trehalose; and negative for adonitol, L-arabitol, D-fucose, L-fucose, inulin, 2-ketogluconate, L-xylose.

* Strain PN-120 unable to grow in the modified API system.

† Strain PN-57 unable to grow in the modified API system.

high NaCl tolerance and ability to deaminate phenylalanine. The morphology of the reference strain showed typical oval endospores in a non-swollen sporangium (Fig. 2a).

Phenon 2. These six bacteria, like those of phenon 3, were obtained from soil or mud samples and had a similar requirement for sodium ions and pH. The mean DNA base composition of these bacteria was lower than that for strains of phenon 3 (39·3 mol% G+C compared with 42·6 mol%; see Table 4). Four of these strains formed a distinct group based on phenotype and DNA hybridization (> 73% DNA reassociation), but DNA could not be isolated from the other two strains, PN-127 and PN-129. The reference strain produced oval spores which clearly distend the sporangium (Fig. 2b). These bacteria can be distinguished by their ability to hydrolyse cellulose and xylan effectively (H. Outtrup, personal communication).

Phenon 3. The three strains in this taxon, all isolated from soil or mud samples, formed a tight cluster at 92% S_j. DNA from the three strains showed a high degree of reassociation and close similarity in mol% G+C composition (Table 4). They were moderately halophilic, showing good growth at 15% NaCl. Indeed, physio-

Table 4. Base composition and hybridization of DNA to reference DNA from strains representing the phenotypic clusters

All figures in parentheses are standard deviations.

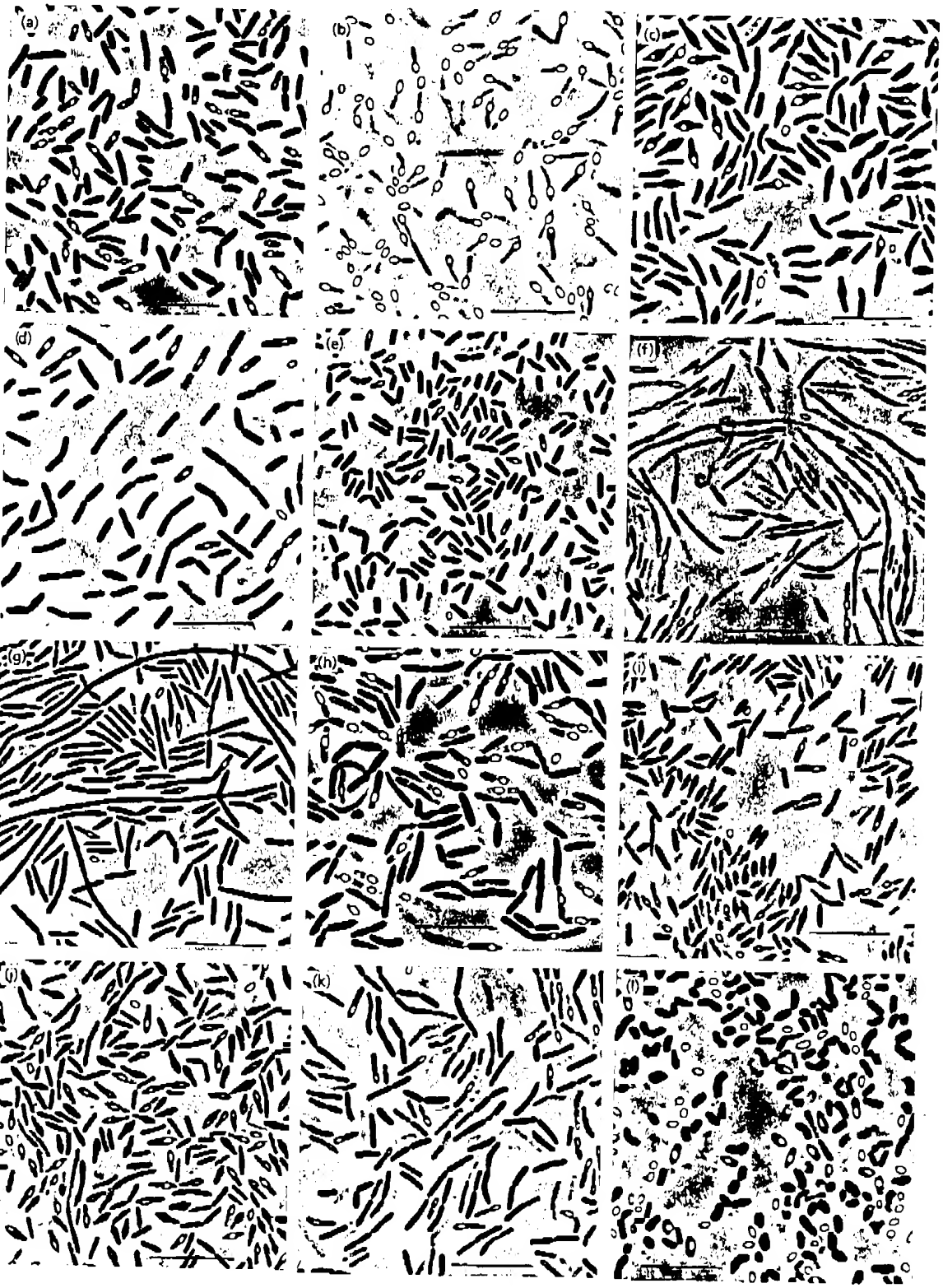
Cluster	Strain†	Mol % G+C	Hybridization to reference DNA (%)										
			1 PN-3	2 PN-105	3 PN-102	4a PN-1	4b PN-26	5 PN-30	6 PN-23	8a PN-118	8b PN-121	9 PN-137	10 DSM 485	11 PN-109
1	18 strains	39·6 (0·4)	95 (16)											
	*PN-3	39·6	100	4	5	12	16	4	15	8	5	5	4	14
	PN-81	39·6	60											
	PN-21	40·2	36											
	PN-73	40·7	31											
	PN-51	40·4	27											
	PN-59	39·1	23											
2	4 strains	39·3 (0·2)		82 (11)										
	*PN-105	39·2	18	100	16	13	13	4	10	9	7	6	4	13
3	3 strains	42·6 (0·2)			93 (9)									
	*PN-102	42·4	9	3	100	13	14	3	11	10	4	7	12	18
4	a: 3 strains	38·6 (0·2)				85 (11)	15 (1)							
	*PN-1	38·4	7	4	7	100	13	7	10	9	6	5	7	12
	b: 5 strains	39·5 (0·2)				17 (5)	87 (9)							
	*PN-26	39·2	6	4	11	23	100	5	19	11	6	8	7	16
	PN-25	38·2				11	11							
5	16 strains	42·6 (0·4)						87 (13)						
	*PN-30	42·8	9	5	11	24	29	100	22	12	8	8	5	19
	PN-17	42·6						60						
	PN-63	42·7						63						
	PN-22	42·6						57						
	PN-44	42·8						51						
	PN-38	42·6						37						
6	6 strains	43·7 (0·8)							78 (10)					
	*PN-23	42·8	5	4	12	19	25	12	100	12	3	7	10	11
	PN-56	43·5							63					
	PN-45	43·2							61					
	PN-75	45·3							13					
8	a: 2 strains	38·6								96	31			
	*PN-118	38·6	20	5	34	15	20	9	39	100	20	10	4	17
	b: 4 strains	41·5 (0·4)								24 (7)	83 (12)			
	*PN-121	41·3	5	4	13	12	14	9	12	18	100	7	4	14
	PN-117	40·6								7	27			
	PN-123	40·6								23	38			
9	4 strains	38·5 (0·3)										81 (11)	13 (3)	
	*PN-137	38·2	16	4	9	11	13	8	10	9	10	100	9	12
	PN-136	39·8										26	17	
10	4 strains	38·2 (0·2)										15 (8)	75 (12)	
	PN-57‡	36·2										6	109	
	*DSM 485	37·0	30	12	16	17	18	5	16	13	15	24	100	25
	PN-100	37·8										9	57	
11	5 strains	41·3 (0·3)												88 (9)
	*PN-109	41·2	26	8	22	30	31	9	30	19	16	19	3	100
	PN-113	41·2												41
Single isolates§	PN-50	34·0	5	5	14	11	12	10	9	10	17	18	14	8

† Strains within clusters are given in Table 1. Strains with more than 65 % sequence homology to the reference strain have been grouped and the mean hybridization value is given. Asterisks show the reactions of the reference strain of the cluster and this strain is included in the group value.

‡ Although strain PN-57 was recovered as a single-membered cluster in Fig. 1 it was included here as a member of cluster 10.

§ These data for single-membered clusters represent a single strain but are typical of the data for all the single-membered clusters shown in Table 1.

logical studies (data not shown) revealed that these bacteria and those of phenon 2 were unique among all the taxa studied in being unable to grow in the absence of sodium ions. They were also unable to grow at pH 7 and showed an optimum for growth at pH 10. A distinctive feature of these bacteria was their inability to hydrolyse



[redacted]

starch. The reference strain produced oval endospores in a clearly swollen sporangium (Fig. 2c).

Phenon 4. This heterogeneous cluster comprised strains allocated to two DNA hybridization groups as well as a single strain which was excluded from these groups (Table 4). The three group 4a strains were largely from soil while the five group 4b strains originated from animal manures. Reference strains from both taxa produced oval endospores which did not swell the sporangia (Fig. 2d, e). The physiological tests and the carbon utilization tests distinguished these taxa but there were no definitive features. Group 4a strains were less temperature tolerant than those of 4b. Utilization of amygdalin, gentiobiose and mannose also provided some distinction between these groups. DNA analysis showed that group 4a strains were slightly lower in G+C content (38·4–38·8 mol %) than group 4b strains (39·2–39·7 mol %).

Phenon 5. The strains in this phenon formed a homogeneous cluster at 90 % S_J which was supported by the DNA hybridization studies, with only five of the 21 strains showing less than 65 % reassociation to the reference strain. '*Bacillus alcalophilus* subsp. *halodurans*' DSM 497 was included in this cluster and therefore its name is used for this group. The slightly swollen sporangia containing oval endospores are shown in Fig. 2(f). The organisms largely originated from animal manure and chicken yard soils and were distinctive in their ability to grow at 55 °C. In spite of the subspecific name *halodurans*, these bacteria had only moderate tolerance to NaCl (around 12 %) in comparison to other groups. However, the name is to be construed in the context of *B. alcalophilus*, which has much lower tolerance to NaCl (Boyer *et al.*, 1973).

Phenon 6. Of the nine strains in phenon 6, three showed less than 65 % DNA hybridization with the reference strain and only one less than 60 %. The long rods typical of the reference strain are shown in Fig. 2(g). These strains had the highest G+C content of all the organisms included in the study (43·7 mol %). They reduced nitrate, were unable to hydrolyse Tween or pullulan, and were fairly sensitive to NaCl.

Phenon 7. This phenon was identified as *B. cohnii* and included several strains from the Wadi Natrun in Egypt and from Mono Lake (California). These strains were excluded from the DNA study because they have been characterized previously (Spanka & Fritze, 1993). Our results confirm the conclusions of the earlier study and emphasize the relatively weak NaCl tolerance of these bacteria.

Phenon 8. The low level at which phenon 8 was defined in the S_J/UPGMA analysis was highlighted by the DNA hybridization study, which revealed two groups of strains and two further strains which were unrelated to these groups but could be related to each other (they share the same base composition; see Table 4). Group 8a (two strains) could be distinguished from group 8b (four strains) by consistent differences in base composition of DNA and the sensitivity of the former strains to 5 % NaCl. Moreover, strains of group 8a, in contrast to most strains of group 8b, failed to hydrolyse Tween 40 and 60 and to grow on glycerol. There were differences in carbohydrate utilization patterns between strains of groups 8a and 8b, particularly with regard to amygdalin and *N*-acetylglucosamine utilization. Morphologically, the reference strains of the two groups were similar but not identical (Fig. 2h, i).

Phenon 9. One of the five phenon 9 strains was unrelated to the reference strain by DNA hybridization and had a slightly higher DNA base composition than other members of the cluster (Table 4). These bacteria failed to grow at pH 7 or above pH 10 and at temperatures above 40 °C. They were tolerant to moderate concentrations of NaCl (10 %). The reference strain produced oval endospores in a slightly swollen sporangium (Fig. 2j).

Phenon 10. This phenon comprised six strains, of which four showed > 74 % DNA reassociation to *B. alcalophilus* DSM 485^T, one showed only 57 % and the sixth did not yield high-quality DNA. The strains in this group showed very similar physiological properties to those of phenon 9 but were less salt tolerant (growth up to 8 % NaCl) and had a slightly lower G+C content. Typical morphology of these cells is shown in Fig. 2(k).

Phenon 11. The six strains in phenon 11 were notable for their homogeneous base composition (41·1 mol % G+C), with only one strain showing less than 65 % DNA hybridization to the reference strain. Phenotypically, they were characterized by lack of starch and pullulan hydrolysis, relatively low tolerance to NaCl and inability to grow at 40 °C or above. The distinctive short fat cells of the reference strain containing oval endospores are shown in Fig. 2(l).

Single-membered clusters. Nine strains were recovered in single-membered clusters and two in a doublet in the S_J/UPGMA analysis. To determine if these bacteria were foci of new clusters or aberrant members of extant clusters, DNA was prepared and hybridized to the DNA from the reference strains of the 11 clusters. Except for strain PN-57, which was allocated to the DNA hybridization group of phenon 10, there was no significant hybridization of the DNA from the single-membered clusters to any of the reference DNAs. This confirmed that the majority of these bacteria did not belong to any of

Fig. 2. Sporulating and non-sporulating cells of representative strains of DNA hybridization groups established in this study. Growth conditions: M1, alkaline nutrient agar supplemented with 5 g NaCl l^{-1} and 10 mg Mn^{2+} ml^{-1}; M2, alkaline nutrient agar supplemented with 5 g NaCl l^{-1} and with 50 % of the water substituted with soil extract. Bars, 10 μm. (a) Strain PN-3 (group 1): M1, 1 d, 30 °C. (b) Strain PN-105 (group 2): M1, 1 d, 30 °C and 4 d, 18 °C. (c) Strain PN-102 (group 3): M1, 2 d, 30 °C. (d) Strain PN-1 (group 4a): M1, 1 d, 30 °C. (e) Strain PN-26 (group 4b): M2, 36 h 30 °C. (f) Strain PN-30 (group 5): M1, 2 d, 30 °C. (g) Strain PN-23 (group 6): M2, 3 d, 25 °C. (h) Strain PN-118 (group 8a): M1, 2 d, 30 °C. (i) Strain PN-121 (group 8b): M1, 36 h, 30 °C. (j) Strain PN-137 (group 9): M1, 1 d, 30 °C. (k) Strain DSM 485 (group 10): M1, 2 d, 30 °C. (l) Strain PN-109 (group 11): M1, 4 d, 30 °C.

the clusters shown in Fig. 1 (data not shown). Moreover, the possibility that these bacteria may be related to each other but failed to cluster in the phenogram was excluded by preparing partial 16S rRNA sequences and analysing them in the context of the full sequences for the reference strains (Nielsen *et al.*, 1994). All of the partial sequences were recovered independently in the phylogenetic tree (data not shown), confirming that they were indeed foci of new species for which we had only single representatives.

DISCUSSION

When faced with a heterogeneous collection of bacteria, the construction of a basic taxonomic framework from which detailed studies of species can be initiated is sometimes a daunting task. Numerical phenetics has become established as a rapid and reliable procedure for generating this initial database in most instances. For example, numerical analysis of *Streptomyces* established the framework for future taxonomic studies of species within this genus (Williams *et al.*, 1983) and a recent numerical analysis of the thermophilic *Bacillus* strains (White *et al.*, 1993) established species which were subsequently evaluated by DNA reassociation and independently by comparative rRNA sequence analysis (Rainey *et al.*, 1994). This current assessment of the diversity of the alkaliphilic *Bacillus* strains confirms numerical phenetics as an ideal approach for gauging the species diversity within a group of bacteria and indicating reference strains for molecular analyses since, with only a few exceptions, the phena defined by the S_J/UPGMA analysis were fully substantiated by DNA reassociation and rRNA sequencing (Nielsen *et al.*, 1994).

When using numerical phenetics in this way, it is a benefit if a minimum of tests can be used and, wherever possible, these can be based on test kits or automated systems. We have confirmed previous numerical analyses (e.g. Priest *et al.*, 1981) which showed that accurate clustering can be detected with fewer than 50 characters, the only problem being that with this reduced database some taxa are not separated. Thus in the above-mentioned study, *B. amyloliquefaciens* could not be distinguished from *B. subtilis* and individual *B. sphaericus* DNA hybridization groups were indistinguishable. Similarly, in this study, clusters 4 and 8 were shown to be heterogeneous by DNA hybridization. Such lack of resolution is to be expected from such a small number of tests (Sneath & Sokal, 1973) and it is likely that with more characters these taxa would have been separated.

API test systems have been used successfully for the classification and identification of *Bacillus* strains (Logan & Berkeley, 1984). However, some alkaliphilic *Bacillus* strains, notably those from clusters 2 and 7, were unable to produce a detectable reaction in the API tests even though the system had been modified for use at high pH. Given the problems of conducting physiological tests at extremes of pH, this was not surprising and illustrates the difficulties involved. For these reasons, in some previous taxonomic studies of these bacteria, the organisms were first 'adapted' to growth at pH 7 and all tests were performed at neutrality (Gordon & Hyde, 1982). However, this is an artificial answer to the problem and may lead to erroneous physiological reactions.

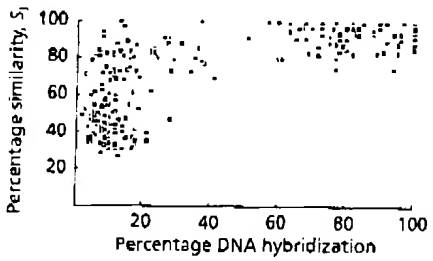


Fig. 3. Comparison of phenotypic similarity calculated from the S_J coefficient with DNA hybridization for 230 pairs of strains.

The numerical analysis was evaluated by determination of the DNA base composition of all the strains in the study and extensive DNA hybridization reactions. In addition to revealing the heterogeneity of clusters 4 and 8, the DNA reassociation reactions also showed that most clusters contained strains which were genomically unrelated, but in all cases these were a minority of strains. We defined DNA hybridization groups at 65%, which is 5% below the currently accepted guideline used in the species definition (Wayne *et al.*, 1987), but given the error rates of around 10% (standard deviation) for our determinations using an immobilized assay, we consider this acceptable. Indeed, this limit was shown to be appropriate by plotting percentage DNA/DNA reassociation against phenetic similarity determined by S_J for 230 pairs of strains (Fig. 3). Although similar results have been presented previously (Staley & Colwell, 1973) they have not been based on so many determinations and do not show so clearly the cut-off of 'species pairs' at a little above 60% DNA hybridization and greater, with most non-homologous pairs showing less than 40% binding. This provides good evidence for the discontinuous spread of DNA hybridization values among these bacterial species and for the use of DNA hybridization for speciation. However, exact definition of the level of DNA hybridization for the species boundary and the difficulties presented by subspecies remain problematic.

It is clear from Fig. 3 that pairs of bacterial strains which show > 65% DNA hybridization invariably show high phenotypic relatedness as judged by S_J and that most strains which show low DNA relatedness present 25–60% phenetic similarity. However, some strains which were phenotypically similar were genomically unrelated. This is probably a result of the limited phenotypic database and reflects the problems associated with heterogeneous clusters. It would be interesting to repeat this exercise with a full numerical taxonomic database, in which case it is likely that these aberrant strains would be correctly placed.

It is also useful to compare the DNA hybridization data with the full 16S rRNA sequence similarities which are

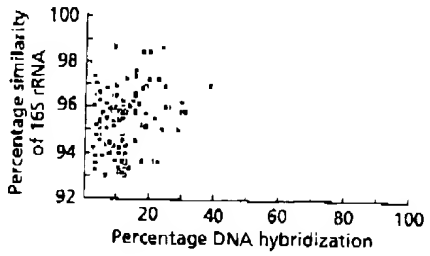


Fig. 4. Comparison of sequence similarity based on the full 16S rRNA sequence (Nielsen *et al.*, 1994) with DNA hybridization for representative alkaliphilic *Bacillus* strains.

available from our previous study (Nielsen *et al.*, 1994). The data in Fig. 4 include only genomically unrelated strains since strains with high DNA hybridization should have virtually identical rRNA sequences. The scatter of points is similar to that obtained by Amann *et al.* (1992) for *Fibrobacter* strains and for *Desulfovibrio* by Devereux *et al.* (1990), and shows that at below approximately 99% rRNA similarity DNA hybridization is below 50% and that strains with rRNA similarities at this level definitely represent distinct genospecies. In contrast to the previous studies, there was no indication that pairs of strains showing rRNA similarity below 94% show lower DNA hybridization than those with rRNA similarity between 94 and 99%. It has been suggested that these lower values could be equated with the generic boundary (Amann *et al.*, 1992).

The phylogenetic positions of all 13 alkaliphilic genospecies which have been defined in this study have been reported by Nielsen *et al.* (1994). Strains which had been included in previous phenetic taxonomic studies were mainly recovered in phena 1, 4a, 4b, 5, 6, 7 and 10, with one new isolate each being assigned to phena 1 and 6 and two to phenon 10. All strains of phena 2, 3, 8a, 8b and 11 were provided by H. Outtrup (Novo Nordisk, Denmark), as were the four new isolates mentioned above, and only limited information from previous studies is available for these strains. Phenon 7 strains were identified as *B. cohnii* and related taxa (Spanka & Fritze, 1993) and are not further discussed; other taxa are discussed separately below.

Phenon 1 (DNA hybridization group 1) contained strains which had previously been identified as *B. firmus* (Gordon & Hyde, 1982), although strain PN-77 was an exception since it had been classified originally as a *B. firmus–B. lentus* intermediate (Gordon & Hyde, 1982). All of these strains, including PN-77, were classified as phenotypic group 1 by Fritze *et al.* (1990). These bacteria are phylogenetically unrelated to *B. firmus* (Nielsen *et al.*, 1994) and their mean DNA base composition is slightly lower than that for *B. firmus* (39·6 mol% against about 41 mol%; Fahmy *et al.*, 1985; Fritze *et al.*, 1990). Moreover, unlike *B. firmus*, most of these bacteria are unable to grow at pH 7, are tolerant of very high levels of NaCl, and deaminate phenylalanine, a variable feature for *B. firmus sensu stricto* (Gordon *et al.*, 1973; Priest *et al.*, 1988). They therefore represent a new species for which we recommend the name '*B. pseudofirmus*'.

Strains in clusters 2 and 3 were phylogenetically removed from all other bacilli. The closest relatives were *B. alcalophilus* DSM 485^T and strains representing clusters 1, 4a, 4b, 5, 6, 9 and 11 of this study (Nielsen *et al.*, 1994). DNA hybridization confirmed that these clusters comprise separate taxa at the species level and they were unique in this study in requiring sodium ions for growth. We suggest the names '*B. agaradhaerens*' and '*B. clarkii*' for clusters 2 and 3 respectively.

Phenon 4 was divided into two hybridization groups but both contained strains which had previously been classified as *B. lentus* type I (Gordon & Hyde 1982) or phenotypic group 4 (Fritze *et al.*, 1990). These bacteria were unrelated to *B. lentus* phylogenetically (Nielsen *et al.*, 1994) and had a higher DNA base composition than *B. lentus* strains (38–39 mol% G+C versus 36 mol% for the type strain of *B. lentus*; Fahmy *et al.*, 1985). Unlike *B. lentus*, these bacteria were proteolytic and grew in 10% NaCl; they cannot therefore be included in this species. Phenon 4 comprised two DNA hybridization groups and the reference strains were unrelated by 16S rRNA sequence analysis (Nielsen *et al.*, 1994). Although these molecular features clearly support two independent and separate taxa, the phenetic data are insufficient to provide diagnostic features at present. The only indications of differing phenotypic properties are the inability of group 4a strains to grow at 40 °C while 4b strains grow at 42 °C, and the slightly larger size of strain PN-1 (4a) cells. To comply with the recommendations of the *Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics* (Wayne *et al.*, 1987), we refrain from naming these taxa until more strains have been isolated and further diagnostic features are available. The proposed reference strain for taxon 4a is PN-1 (= DSM 8714 = NCIMB 10291); that for 4b is PN-26 (= DSM 8717 = NCIMB 10288).

DNA hybridization group 5 comprised 16 strains, including '*B. alcalophilus* subsp. *halodurans*' (Boyer *et al.*, 1973). These strains were previously assigned to *B. lentus* type III (Gordon & Hyde, 1982) and phenotypic group 4 (Fritze *et al.*, 1990). Again, rRNA sequence unrelatedness (Nielsen *et al.*, 1994) and gross difference in mol% G+C (42·6% for phenon 5 strains) exclude these bacteria from *B. lentus* as well as *B. alcalophilus*. We propose to revive the name *halodurans*, in the new combination '*B. halodurans*', for these bacteria.

All strains of DNA hybridization group 6 were previously assigned to *B. lentus* type II by Gordon & Hyde (1982) and to phenotypic group 3 by Fritze *et al.* (1990). These bacteria are distinct from *B. lentus* on the basis of their rRNA sequences (Nielsen *et al.*, 1994), in DNA base composition (43·7 mol% G+C), and in several phenotypic attributes such as casein hydrolysis, nitrate reduction and ability to grow at 50 °C (all negative for *B. lentus* and

positive for DNA hybridization group 6). We therefore propose that these bacteria should be assigned to a new species, for which the name '*B. clausii*' is recommended.

Two independent DNA hybridization groups were defined within phenon 8 strains. Again, the distinct positions of the reference strains according to rRNA sequence analysis underline the validity of these taxa. Although diagnostic phenotypic features are scarce, growth in 5% NaCl and cell morphology (group 8a strains are considerably larger than 8b strains) consistently separate these groups and therefore we propose both as new species: '*B. halmapalus*' for taxon 8a strains and '*B. horikoshii*' for taxon 8b strains.

Clusters 9 and 10 were two further taxa which were difficult to separate phenotypically. Virtually all strains of both phena were recovered in their corresponding DNA hybridization groups and the reference strains had distinct positions according to rRNA sequence analysis. However, the higher NaCl tolerance of group 9 strains compared with those of group 10 and the distinctly swollen sporangia of the former strains provide two easily determined and definitive features for identification. Group 10 strains represent the species *B. alcalophilus* and the data presented here are used to provide an emended description of this previously monotypic species. We propose the name '*B. pseudalcalophilus*' for group 9 strains in recognition of their close phenotypic and phylogenetic relationship with *B. alcalophilus*.

Phenon 11 strains were well defined phenotypically, by DNA hybridization and by RNA sequence analysis (Nielsen *et al.*, 1994). Lack of starch and pullulan hydrolysis and low pH optimum and NaCl tolerance for growth are characteristic of these bacteria. The name '*B. gibsonii*' is proposed for these organisms.

NOMENCLATURE

Emended description of *Bacillus alcalophilus* (Vedder, 1934)

This description is based on the strains of DNA hybridization group 10 of this study. Colonies are white, circular, smooth and shining, sometimes with a darker centre. Cells are rod-shaped (0·5–0·7 × 3·0–5·0 μm), producing ellipsoidal endospores (0·5–0·7 × 0·9–1·3 μm) which are located subterminally and do not swell the sporangium. Strains of this species hydrolyse casein, gelatin, pullulan, starch, Tween 40 and Tween 60. They do not hydrolyse hippurate, MUG or Tween 20 (strain PN-138 is weakly positive). Nitrate is not reduced to nitrite (strain PN-101 is an exception) and strains do not deaminate phenylalanine. No growth is observed at pH 7·0; the optimum is about pH 10 (pH 9·0 for strain PN-57). The temperature range for growth is 10–40 °C. There is a variable reaction to NaCl: the type strain and strain PN-57 are sensitive to 5% NaCl but other strains tolerate 8% NaCl. Carbon source utilization profiles showed that strains use L-arabinose, melibiose and rhamnose for growth. The chromosomal DNA base composition is 36·2–38·4 mol % G + C.

Source: soil and faeces.

Type strain: DSM 485.

Description of *Bacillus agaradhaerens* sp. nov.

Bacillus agaradhaerens (a.gar.ad'hae.rens) sp. nov. Malayan n. *agar* gelling polysaccharide from brown algae; L adj. *adhaerens* adherent; ML adj. *agaradhaerens* adhering to the agar which is characteristic of these colonies.

This description is taken from this paper for DNA hybridization group 2 and unpublished results of H. Outtrup. Colonies are white and rhizoid with a filamentous margin. Cells are rod-shaped (0·5–0·6 × 2·5 μm) and produce ellipsoidal spores (0·6–1·0 × 1·0–1·6 μm) subterminally positioned within a sporangium, which is clearly swollen. Strains of this species are characterized by hydrolysis of Tween 40 and 60 (PN-105 is negative for the former) casein, gelatin, starch, cellulose and xylan. Hippurate, MUG and Tween 20 are not hydrolysed and phenylalanine is not deaminated. Nitrate is reduced to nitrite. Strains of this species are strictly alkaliphilic: no growth is observed at pH 7·0 and optimal growth is observed at pH 10·0 or above. Growth occurs within a temperature range of 10–45 °C. Tolerance to 16% NaCl is observed (strain PN-107 grows only weakly at this NaCl concentration). Carbohydrate utilization profiles show growth on L-arabinose, galactose, mannose, *N*-acetylglucosamine or 2-ketogluconate and most strains grow on tagatose but no growth on methyl β-D-xyloside, inositol or xylitol is observed. The chromosomal DNA composition is between 39·3 and 39·5 mol % G + C.

Source: soil.

Type strain of this species is PN-105 (= DSM 8721).

Description of *Bacillus clarkii* sp. nov.

Bacillus clarkii (clar'ki.i) sp. nov. ML gen. n. *clarkii* from Clark, named after the American bacteriologist Francis E. Clark, who made pioneering studies on the taxonomy of endospore-forming bacteria.

This description is taken from the current study based on strains of DNA hybridization group 3. Colonies are circular with an entire margin and a smooth surface and a cream-white to pale yellow colour (strain PN-104 eventually changes to a dark yellow colour, probably due to synthesis of an excreted pigment). Cells are rod-shaped (0·6–0·7 × 2·0–5·0 μm) and produce ellipsoidal spores (0·6–1·0 × 0·7–1·2 μm) located subterminally. In the type strain, the sporangium is distinctly swollen by the spore. Strains of this species hydrolyse casein, hippurate, gelatin, Tween 40 (strain PN-104 weakly positive) and Tween 60. Starch, MUG, Tween 20 and pullulan are not hydrolysed. Growth is strictly alkaliphilic: no growth is observed at pH 7·0 and optimal growth is above pH 10·0. Growth is observed between 15 and 45 °C and strains are tolerant to 16% NaCl. No growth is observed in minimal medium for carbohydrate utilization at pH 10·0. The chromosomal DNA composition is between 42·4 and 43·0 mol % G + C.

Source: soil.

Type strain of this species is PN-102 (= DSM 8720).

Description of *Bacillus clausii* sp. nov.

Bacillus clausii (clau'si.i) sp. nov. ML gen. n. *clausii* of Claus: named after Dieter Claus, the German bacteriologist who made fundamental contributions to the taxonomy of *Bacillus*.

The description given below is taken from this study for DNA hybridization group 6 strains. Colonies are white and filamentous with a filamentous margin. Cells are rod-shaped (0·5–0·7 × 2·0–4·0 μm), producing spores which are ellipsoidal (0·4–0·6 × 0·7–1·0 μm) and located subterminally to paracentrally in the sporangium, which may be slightly swollen. Cells tend to occur in long chains which may carry spores. The strains in this species are characterized by hydrolysis of casein, gelatin and starch, but not pullulan, Tween 20, 40 or 60 or MUG. Nitrate is reduced but phenylalanine is not deaminated. The pH optimum is about 8, and good growth is obtained at pH 7·0. Growth at temperatures from 15 to 50 °C is observed, with salt tolerance up to about 10% NaCl (strain PN-76 tolerates only 8% NaCl). Carbohydrate utilization patterns show growth on L-arabinose, galactose, xylitol, dulcitol, sorbitol, methyl α-D-mannoside, mannose, *N*-acetylglucosamine, D-tagatose or 2-ketogluconate. Some strains grow on inositol. The chromosomal DNA composition is between 42·8 and 45·5 mol % G + C.

Source: soil.

Type strain of this species is PN-23 (= DSM 8716 = NCIMB 10309).

Description of *Bacillus gibsonii* sp. nov.

Bacillus gibsonii (gib.so'ni.i) sp. nov. J. gen. n. from Gibson, named after the British bacteriologist Thomas Gibson, who made fundamental contributions to the taxonomy of *Bacillus*.

The description is taken from the current study for strains of DNA hybridization group 11. Colonies are yellow and circular with an entire margin and a smooth, shiny surface. Cells are usually stout rods (0·6–1·0 × 2·0–3·0 μm) with ellipsoidal spores (0·6–1·0 × 0·8–1·6 μm) situated subterminally and in ageing cells paracentrally to sometimes lateral. Spores do not cause swelling of the sporangium. Strains of this species hydrolyse casein, gelatin and MUG. Some strains, but not the type strain, hydrolyse Tween 40 and 60. No hydrolysis of starch or pullulan is observed, phenylalanine is not deaminated and reduction of nitrate is variable among strains. Growth occurs at pH 7·0, with an optimum at about pH 8·0. Strains grow in the temperature range from 10 to 37 °C, with the type strain failing to grow at 37 °C. All strains tolerate up to 9% NaCl, while strain PN-111 grows at 12% NaCl. Carbohydrate utilization pattern shows growth on lactose, but not on glycogen or *N*-acetylglucosamine. The chromosomal DNA composition is between 40·6 and 41·7 mol % G + C.

Source: soil.

Type strain of this species is PN-109 (= DSM 8722).

Description of *Bacillus halmapalus* sp. nov.

Bacillus halmapalus (hal.ma'pa.lus) sp. nov. Gr. n. *halme*, brine; Gr. adj. *hapalos*, delicate; ML adj. *halmapalus*, sensitive to brine.

This description is taken from this study based on strains of cluster 8a. Colonies are small, circular with an entire margin, shiny surface and a cream-white colour. Cells are rod shaped (0·6–1·0 × 3·0–4·0 μm) with ellipsoidal spores (0·6–0·8 × 0·9–1·5 μm) located subterminally to paracentrally not swelling the young sporangium. The two strains in this species hydrolyse casein, gelatin, hippurate, pullulan and starch. They do not hydrolyse Tween 20, 40 and 60 or MUG. They grow at pH 7·0, with an optimum at about pH 8·0. Growth is observed between 10 and 40 °C. Salt tolerance is very low, with no growth in the presence of 5% NaCl. Carbohydrate utilization profiles show no growth on glycerol, ribose, D-xylose, L-arabinose, galactose, rhamnose, sorbitol, lactose, melibiose, melizitose, D-raffinose or D-tagatose. The chromosomal DNA composition is 38·6 mol % G + C for the two currently characterized strains.

Source: soil.

Type strain of this species is PN-118 (= DSM 8723).

Description of *Bacillus halodurans* sp. nov.

Bacillus halodurans sp. nov., comb. nov., nom. rev. ('*Bacillus alcalophilus* subsp. *halodurans*', Boyer *et al.*, 1973) ha.lo.du'rans. Gr. n. *hals* salt; L pres. part. *durans* enduring; ML adj. *halodurans* salt-enduring.

This description is taken from this study for strains of DNA hybridization group 5 and Fritze *et al.* (1990) and corresponds largely with the original description. Colonies are white and circular with a slightly filamentous margin. Cells are rod-shaped (0·5–0·6 × 3·0–4·0 μm) and produce spores which are ellipsoidal (0·5–0·6 × 0·8–1·2 μm) and located subterminally in the sporangium, which is slightly swollen. Cells tend to occur in long chains frequently bearing spores. Hydrolysis of Tween 40 and 60, casein, gelatin, starch and pullulan is obtained. Most strains do not hydrolyse Tween 20 (strains PN-62 and PN-80 positive) or hippurate (strain PN-31 positive), and do not reduce nitrate (strains PN-31 and PN-42 positive). MUG is not hydrolysed and phenylalanine is not deaminated. Most strains grow at pH 7·0 (strains PN-31 and PN-42 do not) but optimal growth is obtained around pH 9–10. Growth temperature range is 15–55 °C. Strains show moderate halotolerance; good growth is obtained up to 12% NaCl. Carbohydrate utilization pattern shows growth on L-arabinose, galactose, xylitol, inositol, mannose, *N*-acetylglucosamine or 2-ketogluconate, but not with dulcitol, sorbitol, methyl α-D-mannoside or D-tagatose. The chromosomal DNA composition is between 42·1 and 43·9 mol % G + C.

Source: Soil.

Type strain of this species is PN-80 (= DSM 497).

Description of *Bacillus horikoshii* sp. nov.

Bacillus horikoshii (ho.ri.ko'shi.i) sp. nov. ML gen. n. *horikoshii* of Horikoshi; named after the Japanese microbiologist Koki Horikoshi, who has made fundamental contributions to the study of alkaliphilic bacteria.

This description is taken from this study based on strains of DNA hybridization group 8b. Colonies are small, circular with an entire margin, shiny surface and a cream-white colour. Cells are rod-shaped (0·6–0·7 × 2·0–4·0 μm) with ellipsoidal spores (0·5–0·7 × 0·7–1·2 μm) located subterminally in a sporangium which may be slightly swollen. Strains of this species hydrolyse casein, hippurate, gelatin, pullulan and starch. Three of the four strains hydrolyse Tween 40 and 60. Growth is observed at pH 7·0, with an optimum at about pH 8·0. Strains grow between 10 and 40 °C. Salt tolerance is moderate, with a maximum at 8–9 % NaCl. Strains do not hydrolyse MUG or Tween 20, do not deaminate phenylalanine and do not reduce nitrate to nitrite. No growth is observed on ribose, D-xylose, L-arabinose, galactose, rhamnose, sorbitol, lactose, melibiose, melizitose, D-raffinose or D-tagatose. The chromosomal DNA composition is between 41·1 and 42·0 mol % G+C.

Source: soil samples.

Type strain of this species is PN-121 (= DSM 8719).

Description of *Bacillus pseudalcalophilus* sp. nov.

Bacillus pseudalcalophilus (pseu.dal.ca.lo'phi.lus) sp. nov. Gr. adj. *pseudes* false; specific epithet *alcalophilus*; ML adj. *pseudalcalophilus*, false *alcalophilus* because it is phenotypically closely related to *B. alcalophilus* but phylogenetically distinct.

The description is taken from this study for strains of DNA hybridization group 9. Colonies are white and circular with an undulate margin. Cells are rod shaped (0·5–0·6 × 2·0 4·0 μm) and produce ellipsoidal spores (0·5–0·7 × 0·8–1·3 μm) which swell the sporangium and are situated paracentrally to subterminally. Strains of this species hydrolyse casein, gelatin, starch and pullulan, but hippurate, MUG and Tween 20 are not hydrolysed. Phenylalanine is not deaminated and nitrate is not reduced. No growth is obtained at pH 7·0 and the optimum is about pH 10·0. All strains grow from 10–40 °C. Strains tolerate up to 10 % NaCl. Carbohydrate utilization profiles show growth on L-arabinose or galactose, with some strains able to grow on tagatose or 2-ketogluconate. No growth is observed on inositol, xylitol, dulcitol, sorbitol, methyl α-D-mannoside, *N*-acetylglucosamine or mannose. The chromosomal DNA composition is between 38·2 and 39·0 mol % G+C.

Source: soil samples.

Type strain of this species is PN-137 (= DSM 8725).

Description of *Bacillus pseudofirmus* sp. nov.

Bacillus pseudofirmus (pseu.do.fir'mus) sp. nov. Gr. adj. *pseudes* false; specific epithet *firmus*. ML adj. *pseudofirmus* the false *firmus*; referring to physiological similarities to *B. firmus*.

This description is taken from this study for strains of DNA hybridization group 1. Colonies are yellow and round with irregular margins. Cells are rod-shaped (0·6–0·8 × 3·0–6·0 μm), producing spores which are oval (0·5–0·7 × 0·5–1·2 μm) and located centrally to subterminally without swelling the sporangium. The strains in this species are characterized by hydrolysis of Tween 40 and 60, gelatin, casein and starch. Hippurate, MUG, pullulan (with the exception of PN-3, PN-10 and PN-72) and Tween 20 are not hydrolysed, nitrate is not reduced but all strains deaminate phenylalanine. The strains are strictly alkaliphilic: for most of the strains no growth is obtained at pH 7·0 and the optimum is around pH 9·0. The strains show growth from 10 to 45 °C. Growth is detected in the presence of up to 16 % NaCl, with most strains tolerating 17 % NaCl. The carbohydrate utilization profile shows growth on ribose or D-xylose, but no growth on L-arabinose, galactose, rhamnose, sorbitol, lactose, melibiose, melizitose, D-raffinose or D-tagatose. The chromosomal DNA composition is between 39·0 and 40·8 mol % G+C.

Source: soil and animal manure.

Type strain of this species is PN-3 (= DSM 8715 = NCIMB 10283).

ACKNOWLEDGEMENTS

We would like to thank Helle Outtrup for providing us with bacterial strains which were not publicly available. This work was supported by the Danish Research Academy, DSM – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, and Novo Nordisk. The experimental work was carried out at DSM and at Department of Biological Sciences, Heriot-Watt University. The excellent technical assistance of Katja Weisser is gratefully acknowledged. We thank T. O. Macadoo for assistance with formulation of names.

REFERENCES

Alexander, B. & Priest, F. G. (1989). *Bacillus glucanolyticus*, a new species that degrades a variety of β-glucans. *Int J Syst Bacteriol* **39**, 112–115.

Amann, R. I., Lin, C., Key, R., Montgomery, L. & Stahl, D. A. (1992). Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. *Syst Appl Microbiol* **15**, 23–31.

Ash, C., Farrow, J. A. E., Wallbanks, S. & Collins, M. D. (1991). Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit ribosomal RNA sequences. *Lett Appl Microbiol* **13**, 202–206.

Aunstrup, K., Andresen, O. & Outtrup, H. (1971). UK patent no. 1243784.

Aunstrup, K., Outtrup, H., Andresen, O. & Dambmann, C. (1972). Proteases from alkalophilic *Bacillus* species. In *Fermentation Technology Today; Proceedings of the 4th International Fermentation Symposium*, pp. 299–305. Osaka: Society for Fermentation Technology.

Boyer, E. W., Ingle, M. B. & Mercer, G. D. (1973). *Bacillus alcalophilus*

subsp. *halodurans* subsp. nov.: an alkaline amylase producing alkalophilic organism. *Int J Syst Bacteriol* **23**, 238–242.

Chislett, M. E. & Kushner, D. J. (1961). A strain of *Bacillus circulans* capable of growing under highly alkaline conditions. *J Gen Microbiol* **24**, 187–190.

Demharter, W. & Hensel, R. (1989). *Bacillus thermocloacae* sp. nov., a new thermophilic species from sewage sludge. *Syst Appl Microbiol* **11**, 272–276.

Denariaz, G., Payne, W. J. & Le Gall, J. (1989). A halophilic denitrifier, *Bacillus halodenitrificans* sp. nov. *Int J Syst Bacteriol* **39**, 145–151.

Devereux, R., He, S.-H., Doyle, C. L., Orkland, S., Stahl, D. A., LeGall, J. & Whitman, W. B. (1990). Diversity and origin of *Desulfotibrio* species: phylogenetic definition of a family. *J Bacteriol* **172**, 3609–3619.

Fahmy, F., Flossdorf, J. & Claus, D. (1985). The DNA base composition of type strains of the genus *Bacillus*. *Syst Appl Microbiol* **6**, 60–65.

Fritze, D., Flossdorf, J. & Claus, D. (1990). Taxonomy of alkaliphilic *Bacillus* strains. *Int J Syst Bacteriol* **40**, 92–97.

Gordon, R. E. & Hyde, J. L. (1982). The *Bacillus firmus–Bacillus lentus* complex and the pH 7·0 variants of some alkalophilic strains. *J Gen Microbiol* **128**, 1109–1116.

Gordon, R. E., Haynes, W. C. & Pang, C. H. (1973). The genus *Bacillus* (Agricultural handbook no. 427). Washington, DC: US Department of Agriculture.

Guffanti, A. A. (1983). ATP-dependent Na^+/H^+ antiport activity in *Bacillus alcalophilus* requires generation of an electrochemical gradient of protons. *FEMS Microbiol Lett* **17**, 307–310.

Horikoshi, K. (1971). Production of alkaline enzymes by alkalophilic microorganisms. Part 1. Alkaline protease produced by *Bacillus* no. 221. *Agric Biol Chem* **35**, 1407–1414.

Horikoshi, K. (1975). Process for producing cyclodextrins. US patent no. 3923598.

Horikoshi, K. (1991). *Microorganisms in Alkaline Environments*. Weinheim: VCH Verlagsgesellschaft.

Jones, B. E., Grant, W. D., Collins, N. D. & Mwatha, W. E. (1994). Alkaliphiles: diversity and identification. In *Bacterial Diversity and Systematics*, pp. 195–230. Edited by F. G. Priest, A. Ramos Cormenzana & B. Tindall. New York: Plenum Press.

Johnson, J. L. (1981). Genetic characterization. In *Manual of Methods for General Bacteriology*, pp. 450–472. Edited by G. P. Murray, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg & G. B. Phillips. Washington, DC: American Society for Microbiology.

Kitamoto, N., Kimura, T., Kito, Y. & Ohmiya, K. (1992). Cloning and sequencing of the gene encoding cyclodextrin glucanotransferase from *Bacillus* sp. KL201. *J Ferment Bioeng* **74**, 345–351.

Logan, N. A. & Berkeley, R. C. W. (1984). Identification of *Bacillus* strains using the API system. *J Gen Microbiol* **130**, 1871–1882.

Mesbah, M., Premachandran, U. & Whitman, W. B. (1989). Precise measurement of the G+C content of DNA by HPLC. *Int J Syst Bacteriol* **39**, 159–167.

Nakamura, S., Wakabayashi, K., Nakai, R., Aono, R. & Horikoshi, K. (1993). Production of alkaline xylanase by a newly isolated *Bacillus* sp. YIM-1. *World J Microbiol Biotechnol* **9**, 221–224.

Nielsen, P., Rainey, F. A., Outtrup, H., Priest, F. G. & Fritze, D. (1994). Comparative 16S rDNA sequence analysis of some alkaliphilic bacilli and the establishment of a sixth rRNA group within the genus *Bacillus*. *FEMS Microbiol Lett* **117**, 61–66.

Niimura, Y., Koh, E., Yanagida, F., Suzuki, K.-I., Komagata, K. & Kozaki, M. (1990). *Amphibacillus xylanus* gen. nov., sp. nov., a facultatively anaerobic sporeforming xylan-digesting bacterium which lacks cytochrome, quinone, and catalase. *Int J Syst Bacteriol* **40**, 297–301.

Priest, F. G., Goodfellow, M. & Todd, C. (1981). The genus *Bacillus*: a numerical analysis. In *The Aerobic Endospore-forming Bacteria: Classification and Identification*, pp. 91–103. Edited by R. C. W. Berkeley & M. Goodfellow. London & New York: Academic Press.

Priest, F. G., Goodfellow, M. & Todd, C. (1988). A numerical classification of the genus *Bacillus*. *J Gen Microbiol* **134**, 1847–1882.

Rainey, F. A., Fritze, D. & Stackebrandt, E. (1994). The phylogenetic diversity of thermophilic members of the genus *Bacillus* as revealed by 16S rDNA analysis. *FEMS Microbiol Lett* **115**, 205–211.

Scholz, T., Demharter, W., Hensel, R. & Kandler, O. (1988). *Bacillus pallidus* sp. nov., a new thermophilic species from sewage. *Syst Appl Microbiol* **9**, 91–96.

Seldin, L. & Dubnau, D. (1985). Deoxyribonucleic acid homology among *Bacillus polymyxa*, *Bacillus macerans*, *Bacillus azotofixans*, and other nitrogen-fixing *Bacillus* strains. *Int J Syst Bacteriol* **35**, 151–154.

Skerman, V. B. D., McGowan, V. & Sneath, P. H. A. (editors) (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* **30**, 225–420.

Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. San Francisco: W. H. Freeman.

Spanka, R. & Fritze, D. (1993). *Bacillus cohnii* sp. nov., a new obligately alkaliphilic oval-spore-forming *Bacillus* species with ornithine and aspartic acid instead of diaminopimelic acid in the cell wall. *Int J Syst Bacteriol* **43**, 150–156.

Staley, T. E. & Colwell, R. R. (1973). Application of molecular genetics and numerical taxonomy to the classification of bacteria. *Annu Rev Ecol Syst* **4**, 273–300.

Suzuki, Y., Kishigami, T., Inoue, K., Mizogushi, Y., Eto, N., Takagi, M. & Abe, S. (1983). *Bacillus thermoglucosidasius* sp. nov., a new species of obligately thermophilic Bacilli. *Syst Appl Microbiol* **4**, 487–495.

Tamaoka, J. & Komagata, K. (1984). Determination of DNA base composition by reverse-phase HPLC. *FEMS Microbiol Lett* **25**, 125–128.

Vedder, A. (1934). *Bacillus alcalophilus* n. sp.; benevens enkele ervaringen met sterk alcalische voedingsbodems. *Antonie Leeuwenhoek* **1**, 143–147.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Murray, R. G. E., Stackebrandt, E., Starr, M. P. & Trüper, H. G. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.

White, D., Sharp, R. J. & Priest, F. G. (1993). A polyphasic study of thermophilic bacilli from a wide geographical area. *Antonie Leeuwenhoek* **64**, 357–386.

Williams, S. T., Goodfellow, M., Alderson, G., Wellington, E. M. H., Sneath, P. H. A. & Sackin, M. J. (1983). Numerical classification of *Streptomyces* and related genera. *J Gen Microbiol* **129**, 1743–1813.

Received 23 December 1994; revised 8 March 1995; accepted 23 March 1995.

Nucleic Acids Research, 2000, Vol. 28, No. 21 4317–4331

Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*

Hideto Takami*, Kaoru Nakasone, Yoshihiro Takaki, Go Maeno, Rumie Sasaki, Noriaki Masui, Fumie Fuji, Chie Hirama, Yuka Nakamura, Naotake Ogasawara[1], Satoru Kuhara[2] and Koki Horikoshi

Japan Marine Science and Technology Center, Deep-Sea Microorganisms Research Group, 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan, [1]Nara Institute of Science and Technology, Graduate School of Biological Sciences, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan and [2]Kyushu University, Graduate School of Genetic Resources Technology, Hakozaki, Higashi-ku, Fukuoka 812-8555, Japan

Received June 12, 2000; Revised July 27, 2000; Accepted August 29, 2000

DDBJ/EMBL/GenBank accession nos+

ABSTRACT

The 4 202 353 bp genome of the alkaliphilic bacterium *Bacillus halodurans* C-125 contains 4066 predicted protein coding sequences (CDSs), 2141 (52.7%) of which have functional assignments, 1182 (29%) of which are conserved CDSs with unknown function and 743 (18.3%) of which have no match to any protein database. Among the total CDSs, 8.8% match sequences of proteins found only in *Bacillus subtilis* and 66.7% are widely conserved in comparison with the proteins of various organisms, including *B.subtilis*. The *B.halodurans* genome contains 112 transposase genes, indicating that transposases have played an important evolutionary role in horizontal gene transfer and also in internal genetic rearrangement in the genome. Strain C-125 lacks some of the necessary genes for competence, such as *comS*, *srfA* and *rapC*, supporting the fact that competence has not been demonstrated experimentally in C-125. There is no paralog of *tupA*, encoding teichuronopeptide, which contributes to alkaliphily, in the C-125 genome and an ortholog of *tupA* cannot be found in the *B.subtilis* genome. Out of 11 σ factors which belong to the extracytoplasmic function family, 10 are unique to *B.halodurans*, suggesting that they may have a role in the special mechanism of adaptation to an alkaline environment.

INTRODUCTION

Generally, alkaliphilic *Bacillus* strains cannot grow or grow poorly under neutral pH conditions, but grow well at pH >9.5. Since 1969 we have isolated a great number of alkaliphilic *Bacillus* strains from various environments and have purified many alkaline enzymes (1). Over the past two decades our studies have focused on the enzymology, physiology and molecular genetics of alkaliphilic microorganisms to elucidate their mechanisms of adaptation to alkaline environments. Industrial applications of these microbes have been investigated and some enzymes, such as proteases, amylases, cellulases and xylanases, have been commercialized. It is well recognized that these commercial enzymes have brought great advantages to industry (1). Thus, it is clear that alkaliphilic *Bacillus* strains are quite important and interesting not only academically but also industrially.

An alkaliphilic bacterium, strain C-125 (JCM9153), isolated in 1977, was identified as a member of the genus *Bacillus* and reported as a β-galactosidase (2) and xylanase producer (3). It is the most thoroughly characterized strain, physiologically, biochemically and genetically, among those in our collection of alkaliphilic *Bacillus* isolates (1). Recently, this strain was re-identified as *Bacillus halodurans* based on the results of 16S rDNA sequence and DNA–DNA hybridization analyses (4).

Analysis of the entire genome of *Bacillus subtilis*, which is taxonomically related to alkaliphilic *B.halodurans* strain C-125 (Fig. 1), except for the alkaliphilic phenotype, has been completed (5). Knowledge of the complete nucleotide sequence of the *B.subtilis* genome will definitely facilitate identification of common functions in bacilli and specific functions in alkaliphilic *Bacillus* strains. We have determined the complete genomic sequence of alkaliphilic *B.halodurans* C-125 and its genome is compared with that of *B.subtilis* in this study. Alkaliphilic *B.halodurans* is the second *Bacillus* species whose whole genomic sequence has been completely defined.

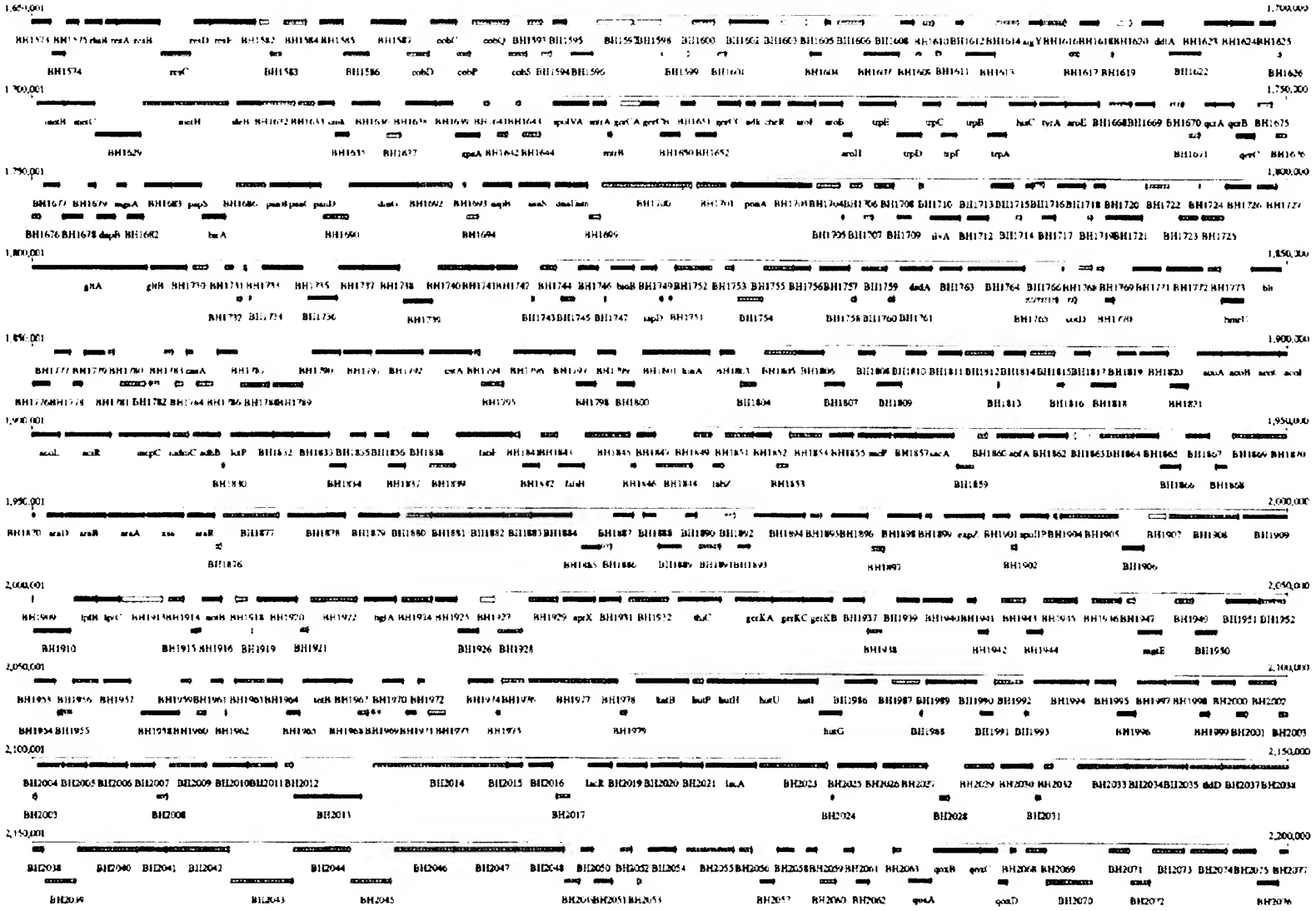
MATERIALS AND METHODS

Preparation of the whole genome shotgun library

A 20 μg aliquot of chromosomal DNA was sonicated for 5–25 s with a Bioruptor UCD-200TM (Tosho Denki Co., Japan). The

*To whom correspondence should be addressed. Tel: +81 468 67 3895; Fax: +81 468 66 6364; Email: takamih@jamstec.go.jp

+AP001507–AP001520



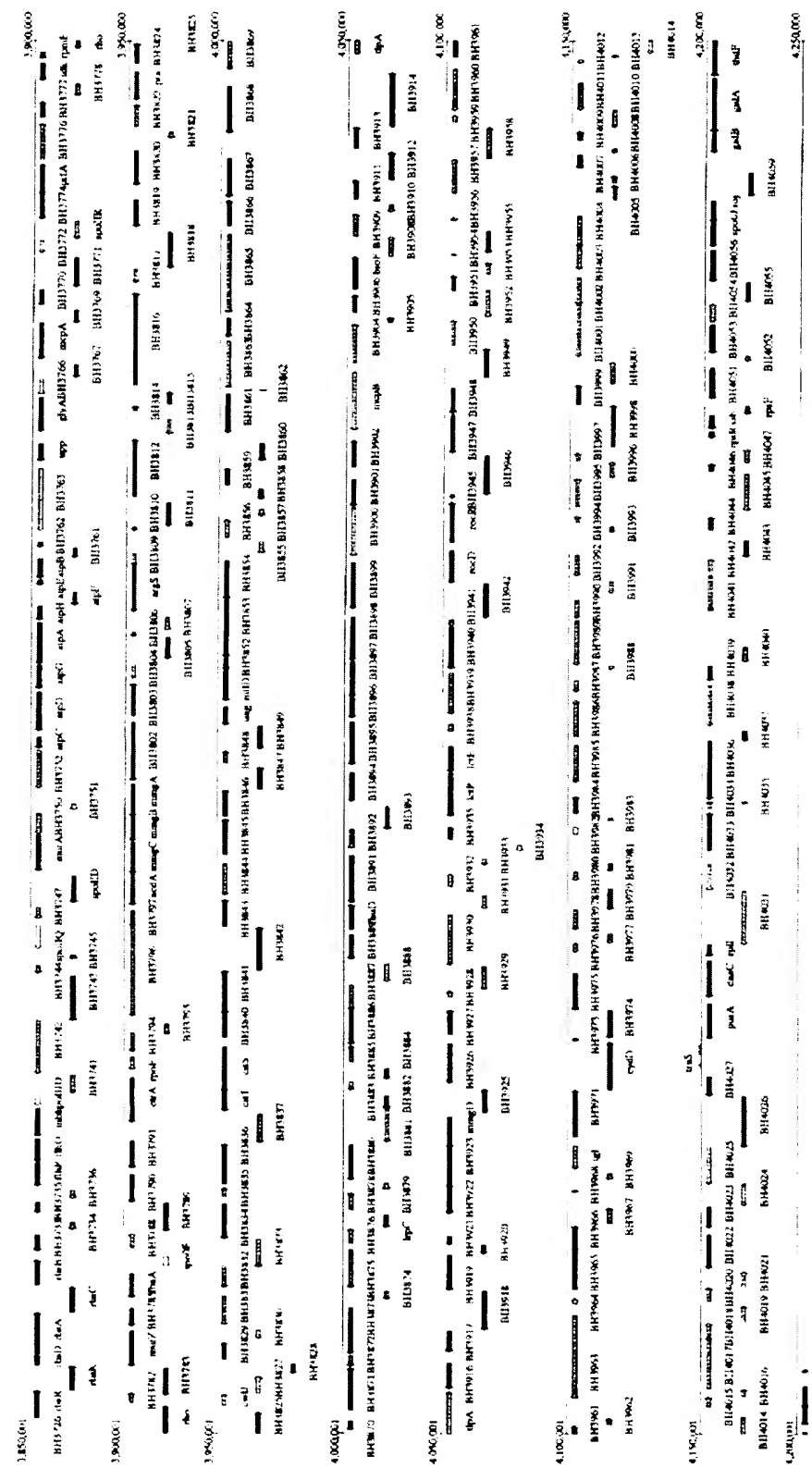


Figure 1. (Above and previous seven pages) General view of the alkaliphilic *B.halodurans* C-125 chromosome. Arrows indicate the direction of transcription. Genes are colored according to their classification into nine broad categories (green, protein conserved in comparison with other organisms, including *B.subtilis*, striped green, unknown, red, protein conserved in comparison with organisms other than *B.subtilis*; striped red, protein conserved in comparison with organisms other than members of *Bacillus*, striped yellow, protein conserved in comparison with other members of *Bacillus*, including *B.subtilis*; striped blue, protein conserved in comparison only with *B.subtilis*, striped pink, protein related to alkaliphily in C-125, purple, protein conserved in comparison with members of *Bacillus* other than *B.subtilis*; see Table S1). Ribosomal RNA genes are colored purple and shown as thin arrows. tRNA genes are marked as vertical bars on the line.

sonicated DNA fragments were blunt-ended using a DNA blunting kit (Takara Shuzo, Kyoto, Japan) and fractionated by 1% agarose gel electrophoresis. DNA fragments 1–2 kb in length were excised from the gel and eluted by the freeze–squeeze method (6). The DNA recovered was ligated to the *Sma*I site of pUC18, which had been previously treated with BAP, and introduced into competent XL1-Blue cells by the standard method (7). Transformants with a frequency of 5–6 × 10^5/μg DNA were cultivated in LB liquid medium at 37°C and 1 μl of culture broth was used for template DNA. The insert in the plasmid, amplified using a standard PCR method, was used for sequencing.

Sequencing and assembly of the whole genome shotgun library

The genome of alkaliphilic *B.halodurans* C-125 was basically sequenced by the whole genome random sequencing method as described (8–10). The DNA fragment inserted into pUC18 was amplified by PCR using M13-20 and reverse primers. PCR fragments, treated with exonuclease I and shrimp alkaline phosphatase (Amersham, OH, USA) to eliminate excess primers in the PCR reaction mixture, were used for sequencing analysis as template DNA. Sequencing was performed with an ABI Prism 377 DNA sequencer using a Taq Dye Terminator Cycle Sequencing Kit (Perkin Elmer, CT, USA). DNA sequences determined by means of the ABI sequencer were assembled into contigs using Phrap (http://bozeman.mbt.washington.edu/phrap.docs/phrap.html) with default parameters and without quality scores. At a statistical coverage of 7.1-fold, the assembly using Phrap yielded 656 contigs. Sequences were obtained from both ends of 2000 randomly chosen clones from a λ library (11). These sequences were then assembled with consensus sequences derived from the contigs of random phase sequences using Phrap. Gaps between contigs were closed by shotgun sequencing of λ clones which bridged the contigs of random phase sequences. The final gaps were closed by direct sequencing of the products amplified by long accurate PCR with a LA PCR Kit v.2 (Takara Shuzo).

Annotation and clustering

The predicted protein coding regions were initially defined by searching for open reading frames (ORFs) longer than 100 codons using the Genome Gambler program (10). Coding potential analysis of the entire genome was performed with the GeneHacker Plus programs using hidden Markov models (12) trained with a set of *B.halodurans* ORFs longer than 300 nt. This program evaluates quality of the Shine–Dalgano sequence (SD) and codon usage for a series of two amino acids. The SD sequence was complementary to one found at the 3′-end of 16S rRNA. The SD sequence (UCUUUCCUCCACUAG...) of alkaliphilic *B.halodurans* C-125 (13) is the same as that of *B.subtilis*. Searches of the protein databases for amino acid similarities were performed using BLAST2 sequence analysis tools (14) with subsequent comparison of protein coding sequences (CDSs) showing significant homology (>10^{-5} significance) performed using the Lipman–Pearson algorithm (15). Significant similarity was defined as at least 30% identity observed over 60% of the CDS, although those CDSs showing <30% identity over >60% of the protein were also included. A search for paralogous gene families, such as σ factors, ATPases, antiporters and ATP-binding cassette (ABC) transporters, in the

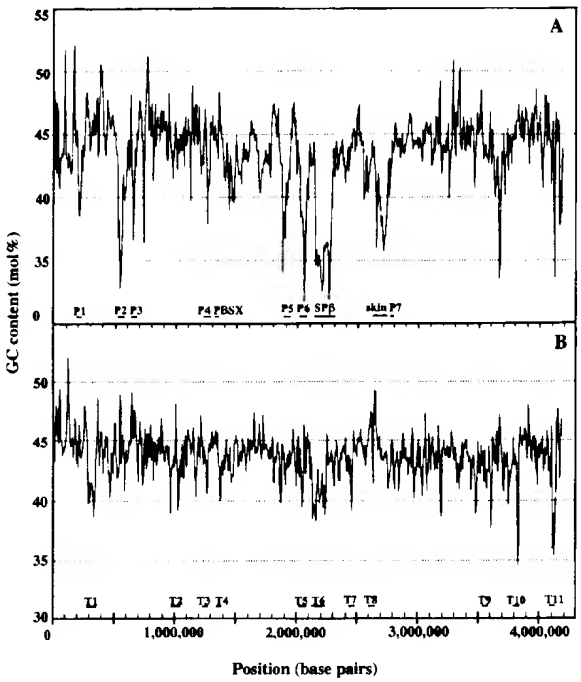


Figure 2. Comparison of G+C profile between the genomes of *B.halodurans* and *B.subtilis*. Distribution of A+T-rich islands along the chromosome, in sliding windows of 10 000 nt, with a step of 5000 nt. Known phage (PBSX, SPβ and skin) are indicated by their names and prophage-like elements in the *B.subtilis* genome are numbered from P1 to P7. The A+T-rich or G+C-rich regions containing transposases in the *B.halodurans* genome are also indicated from T1 to T11. (**A**) *B.subtilis*; (**B**) *B.halodurans*.

B.halodurans genome was performed with stepwise BLAST2P, identifying pairwise matches above $P \leq 10^{-5}$–10^{-80} over 50% of the query search length, and subsequently by single linkage clustering of these matches into multigene families (16).

RESULTS AND DISCUSSION

Genome analysis

The genome of *B.halodurans* is a single circular chromosome (17) consisting of 4 202 353 bp (Fig. 1) with an average G+C content of 43.7% (Table 1). The G+C content of DNA in the coding regions and non-coding regions is 44.4 and 39.8%, respectively. On the basis of analysis of the G+C ratio and G–C skew (G–C/G+C), we estimated that the site of termination of replication (*terC*) is nearly 2.2–2.3 Mb (193°) from the origin, but we could not find the gene encoding the replication termination protein (*rtp*) in the genome of *B.halodurans*. Several A+T-rich and G+C-rich islands are likely to reveal the signature of transposons or other inserted elements (Fig. 2). We identified 4066 CDSs (Fig. 1 and Supplementary Table S1 available at NAR Online), on average 877 nt in size, using the coding region analysis program GeneHacker Plus (12) and the Genome Gambler system (10). We have not annotated CDSs that largely or entirely overlap existing genes. It was found that the termination codon in BH1054, annotated as a transposase,

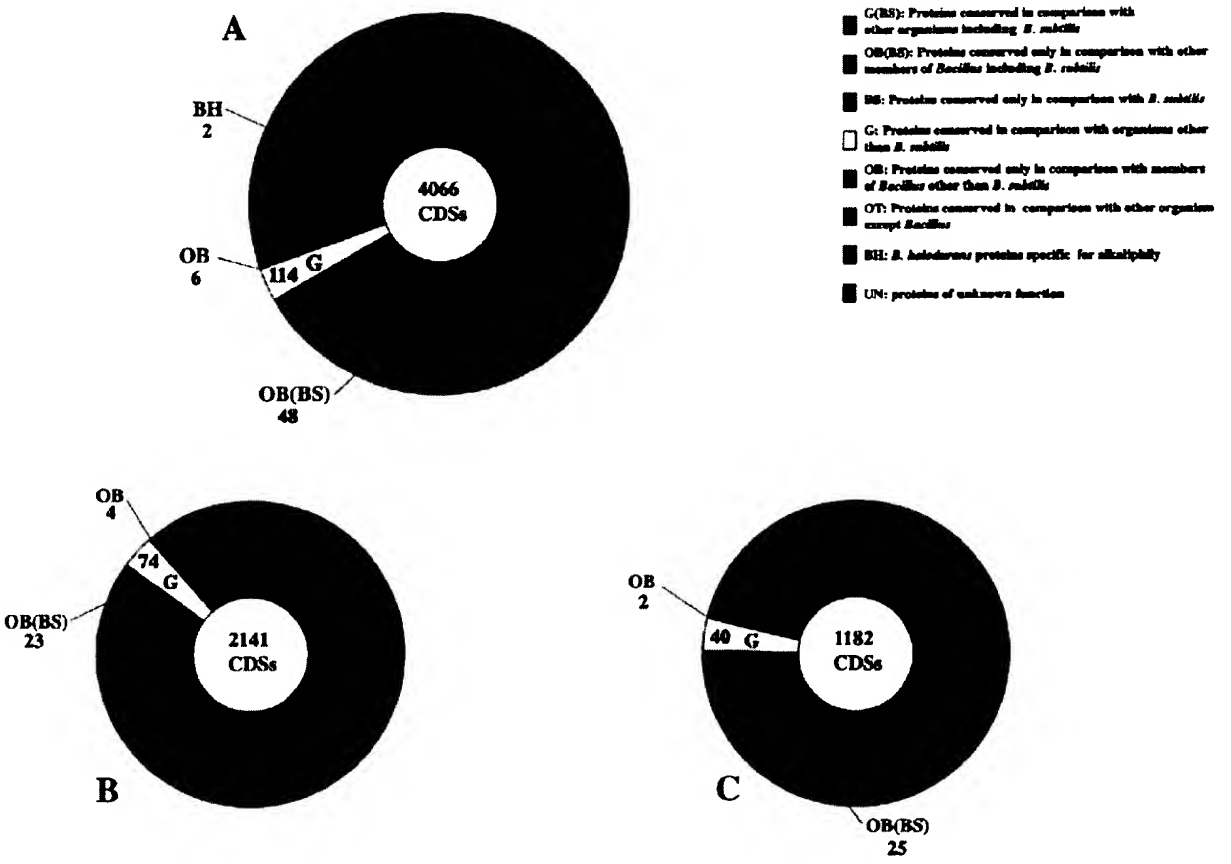


Figure 3. Summary of conserved CDSs identified in the genome of *B.halodurans* C-125. (**A**) All CDSs; (**B**) CDSs with a known functional role; (**C**) CDSs of unknown function

disappeared due to a frameshift and as a result this CDS was combined with the adjacent CDS, presumably coding for an ABC transporter/ATP-binding protein. We identify it as a gene coding for a transposase in this case. Coding sequences cover 85% of the chromosome. We found that 78% of the genes started with ATG, 10% with TTG and 12% with GTG, as compared with 87, 13 and 9%, respectively, in the case of *B.subtilis* (Table 1). The average size of the predicted proteins in *B.halodurans* is 32.841 kDa, ranging from 1.188 to 199.106 kDa. Predicted protein sequences were compared with sequences in a non-redundant protein database and biological roles were assigned to 2141 (52.7%) of them. In this database search 1182 predicted coding sequences (29.1%) were identified as conserved proteins of unknown function in comparison with proteins from other organisms, including *B.subtilis*, and for 743 (18.3%) there was no database match (Table 1). Among all of the CDSs found in the *B.halodurans* genome, 2310 (56.8%) were widely conserved in other organisms, including *B.subtilis*, and 355 (8.7%) of the CDSs matched the sequences of proteins found only in *B.subtilis* (Fig. 3). The ratio of proteins conserved in various organisms, including *B.subtilis*, among functionally assigned CDSs (2141) and among the CDSs (1182) matched with hypothetical proteins from other organisms was 80.5 and 49.7%, respectively, as shown in Figure 3. Of 1183 CDSs, 23.8% matched hypothetical proteins found only in the *B.subtilis* database, showing relatively high similarity values (Fig. 3).

General features

Bacillus halodurans C-125 is quite similar to *B.subtilis* in terms of genome size, G+C content of the genomic DNA and the physiological properties used for taxonomical identification, except for the alkaliphilic phenotype (4). Also, the phylogenetic placement of C-125 based on 16S rDNA sequence analysis indicates that this organism is more closely related to *B.subtilis* than to other members of the genus *Bacillus*. Therefore, the question arises of how does the genome structure differ between two *Bacillus* strains which have similar properties except for alkaliphily. As a first step to answer this question, we analyzed the genome structure both at the level of the whole genomic sequence and at the level of orthologous proteins, comparing the *B.halodurans* and *B.subtilis* genomes continuously from the replication origin region (*oriC*). The dots in Figure 4A were plotted when more than 20 bases in the *B.halodurans* nucleotide sequence continuously matched those of *B.subtilis* in a sliding window 100 nt wide, with a step of

50 nt. Figure 4B shows the distribution of orthologous proteins, comparing *B.halodurans* and *B.subtilis*, and the dot patterns in these figures resemble each other. About 1500 genes, some of which constitute operons, mainly categorized as genes associated with the following functions, are well conserved in the region common to *B.halodurans* and *B.subtilis*: mobility and chemotaxis, protein secretion, cell division, the main glycolytic pathways, the TCA cycle, metabolism of nucleotides and nucleic acids, metabolism of coenzymes and prosthetic groups, DNA replication, RNA modification, ribosomal proteins, aminoacyl-tRNA synthetases, protein folding, etc. On the other hand, the region around 112–153° in the *B.halodurans* genome corresponds to the region around 212–240° in the *B.subtilis* genome, as suggested in a previous paper (17).

Transposable elements

One hundred and twelve CDSs in the *B.halodurans* genome showed significant similarity to the transposases or recombinases from various species, such as *Anabena* sp. *Rhodobacter capsulatus*, *Lactococcus lactis*, *Enterococcus faecium*, *Clostridium beijerinckii*, *Staphylococcus aureus* and *Yersinia pseudotuberculosis*, indicating that these have played an important evolutionary role in horizontal gene transfer and also in internal rearrangement of the genome. These CDSs were categorized into 27 groups by similarity pattern and the genes are widely spread throughout the genome (Fig. 1 and Table S1). As shown in Figure 2, at least 11 A+T-rich and G+C-rich islands containing tranposases (T1–T11) occur in the *B.halodurans* genome. This is one of the notable features of this genome, as *B.subtilis* has only 10 transposons and transposon-related proteins. The G+C content of transposases varies from 37.4 to 49.2% and codon usage in transposases, especially for termination, is obviously different from other indigenous genes in *B.halodurans*. In other bacterial genomes it has been reported that *Synecosystis* sp. PC6803 (18), *Escherichia coli* MG1655 (19), *Mycobacterium tuberculosis* (20), *Deinicoccus radiodurans* (21) and *Lactococcus lactis* (22) contain many transposase genes, as well as *B.halodurans* C-125.

Table 1. Comparison of the general features of the *B.halodurans* and *B.subtilis* genomes

Genome features	*B.halodurans*	*B.subtilis*[a]
Size (bp)	4 202 353	4 214 810
G+C content (mol%)		
Total genome	43.7	43.5
Coding region	44.4	44.3
Non-coding region	39.8	39.3
Open reading frames		
Percent of genome (coding)	85	87
Predicted no.	4066	4100
Conserved with function assigned	2144	2379
Conserved with unknown function	1182	668
Non-conserved	743	1053
Percent AUG initiation codons	78	78
Percent GUG initiation codons	12	9
Percent UUG initiation codons	10	13
Insertion elements		
PBSX prophage-related protein	2	27
Transposase and related protein	112	10
RNA elements		
Stable RNA (percent of genome)	1.02	1.27
16S, 23S and 5S rRNA	8	10
tRNA	78	86

[a]Kunst *et al.* (5).

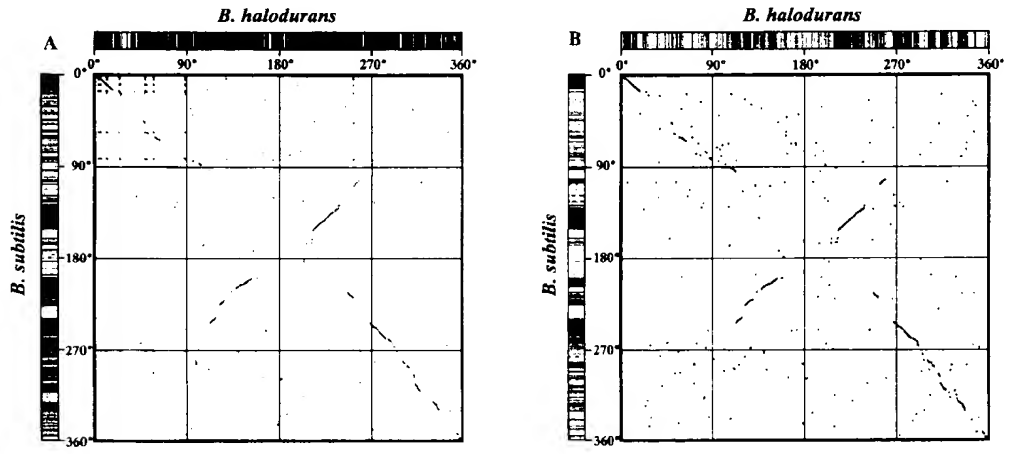


Figure 4. Structural analysis of the *B.halodurans* and *B.subtilis* genomes. (**A**) Distribution of regions having a nucleotide sequence common to the two *Bacillus* spp. The dots were plotted when more than 20 bases of the nucleotide sequence matched continuously with that of *B.subtilis* in a sliding window 100 nt wide with a step of 50 nt. The columns appearing as a bar code clearly show the regions shared between the *B.halodurans* and *B.subtilis* genomes. (**B**) Distribution of orthologs between the two *Bacillus* spp. The dots were plotted at the positions where the gene for an orthologous protein exists when comparing the two genomes. The columns appearing as a bar code clearly show the positions of the genes for orthologous proteins in *B.halodurans* and *B.subtilis*.

On the other hand, we could not find any prophage which seemed to be active, although several phage-related proteins were identified in the *B.halodurans* genome. The *B.halodurans* genome contains no intact prophage, such as Spβ, PBSX or skin, found in the *B.subtilis* genome (5), as shown in Figure 2. We confirmed that *B.halodurans* has the gene for σ^K as the complete form, which is divided into two parts, *spoIVCB* (N-terminal) and *spoIIIC* (C-terminal) by a prophage (skin element) in the *B.subtilis* genome.

Origin of replication

There are 14 CDSs in the *oriC* region of the chromosome of *B.halodurans*. The organization of the CDSs in the region is basically similar to those of other bacteria. The region from *gidB* to *gyrA* (BH4060–BH4066 and BH1–BH7), especially, was found to be the same as in *B.subtilis*. On the other hand, it was found that there are 10 CDSs (BH8–BH18), including three CDSs previously identified in the 13.3 kb of the *oriC* region (23), between *gyrB* and the *rrnA* operon, corresponding to the *rrnO* operon in the *B.subtilis* genome (Fig. 1 and Table S1), although there is no CDS between *gyrA* and *rrnO* in *B.subtilis*. Of these 10 CDSs, only one (BH8) was found to have a homolog in another organism, interestingly, not in the genus *Bacillus*; the others were unique to the *B.halodurans* genome (Table S1).

Transcription and translation

Genes encoding the three subunits (α, β and β′) of the core RNA polymerase have been identified in *B.subtilis* along with the genes for 19 σ factors (24). σ factors belonging to the σ^{70} family (σ^A, σ^B, σ^D, σ^E, σ^F, σ^G, σ^H and σ^K) required for sporulation and σ^L are well conserved between *B.halodurans* and *B.subtilis*. Of 11 σ factors identified in *B.halodurans*, belonging to the extracytoplasmic function (ECF) family, σ^W is also found in *B.subtilis*, but the other 10 (BH620, BH672, BH1615, BH2026, BH3117, BH3216, BH3223, BH3362, BH3632 and BH3882) are unique to *B.halodurans*. These unique σ factors may have a role in the special physiological mechanisms by which *B.halodurans* is able to live in an alkaline environment, because it is well known that ECF σ factors are present in a wide variety of bacteria and they serve to control the uptake or secretion of specific molecules or ions and to control responses to a variety of extracellular stress signals (25).

Seventy-nine tRNA species, organized into 11 clusters involving 71 genes plus eight single genes, were identified (Fig. 1 and Table S1). Of the 11 clusters, six were organized in association with rRNA operons. Eight rRNA operons are present in the C-125 genome and their organization is the same as that in *B.subtilis* (tRNA–16S–23S–5S, 16S–tRNA–23S–5S and 16S–23S–5S–tRNA). With respect to tRNA synthetases, the C-125 genome lacks the glutaminyl-tRNA synthetase gene (*glnS*), one of two threonyl-tRNA synthetase genes (*thrZ*) and one of two tyrosyl-tRNA synthetase genes (*tyrS*). The *B.subtilis* genome has all of these tRNA synthetase genes except for the glutaminyl-tRNA gene. It is likely that glutamyl-tRNA synthetase aminoacylates $tRNA^{Gln}$ with glutamate followed by transamidation by Glu-tRNA amidotransferase (26) in both *Bacillus* species.

Competence and sporulation

Out of 20 genes related to competence in *B.subtilis*, 13 (*cinA*, *comC*, *comEA*, *comEB*, *comEC*, *comER*, *comFA*, *comFC*, *comGA*, *comGB*, *comGC*, *comGD* and *mecA*), mainly expressed in the late stage of competence, were identified in the *B.halodurans* genome, but we could not find any of the genes expressed in the early stage of competence. Among six genes whose products are known to serve as components of the DNA transport machinery, only three (*comGB*, *comGC* and *comGD*), but not the others well conserved in *B.subtilis*, were identified in *B.halodurans* C-125. Actually, competence has not been experimentally demonstrated in C-125 although we attempted to use standard (27) and modified methods, changing such conditions as pH, temperature and medium, for transformation. It has become clear that this is due to lack of some of the necessary genes, especially those expressed in the early stages, such as *comS*, *srfA* and *rapC*. Only 68 genes related to sporulation were identified in the C-125 genome, in contrast to 138 genes found in the *B.subtilis* genome. Although the minimum set of genes for sporulation was well conserved, as in the case of *B.subtilis*, the C-125 genome lacks some genes encoding key regulatory proteins (the response regulator for aspartate phosphatase and the phosphatase regulator) and the spore coat protein for sporulation conserved in the *B.subtilis* genome. The *rap* (*rapA–rapK*) and *phr* (*phrA*, *phrC*, *phrE–phrG*, *phrI* and *phrK*) genes especially were not found in the C-125 genome, suggesting that C-125 may have another type(s) of regulatory gene(s) for control of sporulation in a manner the same as or different from that in *B.subtilis*, as we have observed sporulation in *B.halodurans*.

Cell walls

The peptidoglycan of alkaliphilic *B.halodurans* C-125 appears to be similar to that of neutrophilic *B.subtilis*. However, the cell wall components in C-125 are characterized by an excess of hexosamines and amino acids compared to that of *B.subtilis*. Glucosamine, muramic acid, D- and L-alanine, D-glutamic acid, *meso*-diaminopimelic acid and acetic acid were found in cell wall hydrolysates (1). Although some variation was found in the amide content of the peptidoglycan isolated from alkaliphilic *B.halodurans* C-125, the pattern of variation was similar to that known to occur in *B.subtilis*. All genes related to peptidoglycan biosynthesis, such as *mraY*, *murC–murG*, *cwlA*, *ddlA* and *glnA*, confirmed to be present in the *B.subtilis* genome were also conserved in the C-125 genome (1). A bacitracin resistance gene found in the *B.subtilis* genome is duplicated in the C-125 genome (BH464 and BH1521). On the other hand, although the *tagH* and *tagG* genes were identified in *B.halodurans* C-125 (Fig. 1 and Table S1), 13 other genes for teichoic acid biosynthesis found in *B.subtilis* (*dltA–dltE*, *ggaA*, *ggaB*, *tagA–tagC*, *tagE*, *tagF* and *tagO*) are missing from the *B.halodurans* genome. *Bacillus halodurans* also lacks six genes (*tuaB–tuaF* and *tuaH*) for teichuronic acid biosynthesis, all except *tuaA* and *tuaG*, in comparison with those of *B.subtilis*. In addition to peptidoglycan, the cell wall of alkaliphilic *B.halodurans* is known to contain certain acidic polymers, such as galacturonic acid, glutamic acid, aspartic acid and phosphoric acid. A teichuronopeptide (TUP) is present as a major structural component of the cell wall of C-125, which is a co-polymer of polyglutamic acid and polyglucuronic acid.

Thus, the negative charges on acidic non-peptidoglycan components may give the cell surface the ability to absorb sodium and hydronium ions and to repel hydroxide ions and, as a consequence, may contribute to allowing the cells to grow in alkaline environments. A mutant defective in TUP synthesis grows slowly at alkaline pH. The upper limit of pH for growth of the mutant is 10.4, whereas that of the parental C-125 strain is 10.8. The *tupA* gene encoding TUP has been cloned from the C-125 chromosomal DNA (28). In this study, it has become clear that *B.halodurans* C-125 has no paralog of *tupA* in the genome and an ortholog of *tupA* cannot be found in the *B.subtilis* genome.

Membrane transport and energy generation

Bacillus halodurans C-125 requires Na^+ for growth under alkaline conditions. The presence of sodium ions in the surrounding environment has been proved to be essential for effective solute transport through the cytoplasmic membrane of C-125 cells. According to the chemi-osmotic theory, a proton-motive force is generated across the cytoplasmic membrane by the electron transport chain or by extrusion of H^+ derived from ATP metabolism through the action of ATPase. We identified four types of ATPases (preprotein translocase subunit, class III heat shock ATP-dependent protease, heavy metal transporting ATPase and cation transporting ATPase). These ATPases are well conserved between *B.halodurans* and *B.subtilis*.

Through a series of analyses such as a BLAST2 search, clustering analysis by the single linkage method examining all CDSs identified in the *B.halodurans* C-125 and *B.subtilis* genomes (8166 CDSs) and multiple alignment (16), 18 CDSs were grouped into the category of antiporter- and transporter-related protein genes in the C-125 genome. In this analysis it was found that five CDSs are candidates for Na^+/H^+ antiporter genes (BH1316, BH1319, BH2844, BH2964 and BH3946). However, we could not find any gene encoding antibiotic resistance proteins in the C-125 genome, whereas the *B.subtilis* genome has nine different ones. Eleven genes for multidrug resistance proteins were identified in the C-125 genome, six fewer than in *B.subtilis*. A non-alkaliphilic mutant strain (mutant 38154) derived from *B.halodurans* C-125 which is useful as a host for cloning genes related to alkaliphily has been isolated and characterized (29). A 3.7 kb DNA fragment (pALK fragment) from the parent strain restored growth of mutant 38154 under alkaline pH conditions. This fragment was found to contain CDS BH1319, which is one of the Na^+/H^+ antiporter genes in *B.halodurans*. The transformant was able to maintain an intracellular pH lower than the external pH and the cells expressed an electrogenic Na^+/H^+ antiporter driven only by $\Delta\psi$ (membrane potential, interior negative) (1,29). *Bacillus subtilis* has an ortholog (*mprA*) of BH1319 and it has been reported that a *mprA*-deficient mutant of *B.subtilis* showed a sodium-sensitive phenotype (30). On the other hand, a mutant of strain C-125 with a mutation in BH1317 adjacent to BH1319 has been isolated and it showed an alkali-sensitive phenotype, although whether the Na^+/H^+ antiporter encoded by BH1317 is active in this mutant has not been confirmed experimentally yet. In addition, it has been reported that BH2819, the function of which is unknown and which is unique to the C-125 genome, is also related to the alkaliphilic phenotype (31).

Bacillus halodurans C-125 has a respiratory electron transport chain and the basic gene set for it is conserved as compared with *B.subtilis*, but the gene for cytochrome *bd* oxidase (BH3775 and BH3776) is duplicated in the C-125 genome. It is also clear that two genes for *bo*3-type cytochrome *c* oxidase (BH739 and BH740) not seen in *B.subtilis* are present in the C-125 genome. The C-125 genome has a F_1F_0-ATP synthase operon (Fig. 1 and Table S1). The gene order in this operon (ε subunit–β subunit–γ subunit–α subunit–δ subunit–subunit b–subunit c–subunit a) is identical to that seen in *B.subtilis*. In addition to the F_1F_0-ATP synthase operon, the operon for a Na^+-transporting ATP synthase and the operon for a flagellar-specific ATP synthase are also conserved between *B.halodurans* and *B.subtilis*.

ABC transporters

Members of the superfamily of ABC transport systems couple the hydrolysis of ATP to the translocation of solutes across a biological membrane (32). ABC transporter genes are the most frequent class of protein coding genes found in the *B.halodurans* genome, as in the case of *B.subtilis*. They must be extremely important in Gram-positive bacteria such as *Bacillus*, because these bacteria have an envelope consisting of a single membrane. ABC transporters allow such bacteria to escape the toxic action of many compounds. Through the series of analyses described above, 75 genes coding for ABC transporter/ATP-binding proteins were identified in the *B.halodurans* genome. In this analysis 67 CDSs were grouped into the category of ATP-binding protein genes, although 71 ATP-binding protein genes have been identified in the *B.subtilis* genome (5). We found that *B.halodurans* has eight more oligopeptide ATP-binding proteins, but four fewer amino acid ATP-binding proteins, as compared with *B.subtilis*. We could not find any other substantial difference between *B.halodurans* and *B.subtilis* in terms of the other ATP-binding proteins, although it should be noted that the specificity of some of these proteins is not known. The genes for oligopeptide ATP-binding proteins (BH27, BH28, BH570, BH571, BH1799, BH1800, BH2077, BH2078, BH3639, BH3640, BH3645, BH3646, AppD and AppF) are distributed throughout the C-125 genome. We speculate that these may contribute to survival under highly alkaline conditions, although there is no direct evidence to support this. On the other hand, 43 CDSs were identified as ABC transporter/permeases in the *B.halodurans* genome. Surprisingly, *B.halodurans* has only one amino acid permease, in contrast to the 12 present in the *B.subtilis* genome. In addition, it is clear that *B.halodurans* lacks the sodium permease gene present in *B.subtilis*, whereas *B.subtilis* lacks the nickel permease gene present in *B.halodurans*.

CONCLUSION

Alkaliphilic *B.halodurans* is the second *Bacillus* species whose whole genomic sequence has been completely defined. The genomic sequence of *B.halodurans* offers a wealth of basic information regarding gene conservation and diversity in *Bacillus* spp. and systematic information that would be difficult, if not impossible, to obtain by any other approach. A more complete understanding of the biochemistry of this organism derived from genome analysis will provide the

foundation for clarification of the mechanisms of adaptation to extreme environments, especially to a highly alkaline environment, as a first step. A new database specifically established for the *B.halodurans* sequence, ExtremoBase, will be accessible through the World Wide Web server at http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html . The sequence has been deposited in DDBJ/EMBL/GenBank with the accession nos AP001507–AP001520.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to A. Ohyama, K. Doga, and T. Kozuki of Mitsui Knowledge Inc. Ltd for computing assistance. We thank Dr T. Yada of RIKEN Genomic Science Centre for his help with operating the GeneHacker Plus program. Thanks are also due to Dr T. Sakiyama for his technical assistance.

REFERENCES

1. Horikoshi,K. (1999) *Microbiol. Rev.*, **63**, 735–750.
2. Ikura,Y. and Horikoshi,K. (1979) *Agric. Biol. Chem.*, **43**, 85–88.
3. Honda,H., Kudo,T., Ikura,Y. and Horikoshi,K. (1985) *Can. J. Microbiol.*, **31**, 538–542.
4. Takami,H. and Horikoshi,K. (1999) *Biosci. Biotechnol. Biochem.*, **63**, 943–945.
5. Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S. *et al.* (1997) *Nature*, **390**, 249–256.
6. Thuring,R.W., Sanders,J.M.P. and Borst,P. (1975) *Anal. Biochem.*, **66**, 213–220.
7. Sambrook,J., Frisch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
8. Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A., Merrick,J.M. *et al.* (1995) *Science*, **269**, 496–512.
9. Takami,H. and Horikoshi,K. (2000) *Extremophiles*, **4**, 99–108.
10. Sakiyama,T., Takami,H., Ogasawara,N., Kuhara,S., Doga,K., Ohyama,A. and Horikoshi,K. (2000) *Biosci. Biotechnol. Biochem.*, **64**, 670–673.
11. Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y., Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K. (1999) *Extremophiles*, **3**, 29–34.
12. Yada,T. and Hirosawa,M. (1996) In *Proceedings of the Fourth International Conference on Intelligent Systems and Molecular Biology*. AAAI Press, St Louis, MO, pp. 252–260.
13. Takami,H., Inoue,A., Fuji,F. and Horikoshi,K. (1997) *FEMS Microbiol Lett.*, **152**, 279–285
14. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.
15. Pearson,W.R. and Lipman,D.J (1988) *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
16. Ogiwara,A. (1998) *Genome Informatics*, **1998**, 310–311.
17. Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fuji,F., Nakamura,Y. and Inoue,A. (1999) *Extremophiles*, **3**, 21–28.
18. Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S. *et al.* (1996) *DNA Res.*, **3**, 109–136.
19. Blattner,F.R., Plunkett,G.III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G F. *et al.* (1997) *Science*, **277**, 1453–1474.
20. Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry,C.E.III *et al.* (1998) *Nature*, **393**, 537–544.
21. White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L. *et al* (1999) *Science*, **286**, 1571–1577.
22. Bolotin,A., Mauger,S., Malarme,K., Ehrlich,S.D. and Sorokin,A. (1999) *Antonie Van Leeuwenhoek*, **76**, 27–76.
23. Takami,H., Masui,N., Nakasone,K. and Horikoshi,K. (1999) *Biosci. Biotechnol. Biochem.*, **63**, 1134–1137.
24. Haldenwang,W.G. (1995) *Microbiol. Rev.*, **59**, 1–30.
25. Lonetto,M.A., Brown,K.L., Rudd,K.E. and Buttner,M.J. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 7573–7577.
26. Freist,W., Gauss,D.H., Ibba,M and Soll,D. (1997) *Biol. Chem.*, **378**, 1103–1117.
27. Dubnau,D. and Davidoff-Abelson,R. (1971) *J. Mol. Biol.*, **56**, 209–221
28. Aono,R., Ito,M. and Machida,T. (1999) *J. Bacteriol.*, **181**, 6600–6606.
29. Kudo,T., Hino,M., Kitada,M. and Horikoshi,K. (1990) *J. Bacteriol.*, **172**, 7282–7283.
30. Kosono,S., Ohhashi,Y., Kawamura,F., Kitada,M. and Kudo,T. (2000) *J. Bacteriol.*, **182**, 898–904.
31. Schneider,E. and Hunke,S. (1998) *FEMS Microbiol. Rev.*, **22**, 1–20.
32. Aono,R., Hashimoto,M., Hayakawa,A., Nakamura,S. and Horikoshi,K. (1992) *Biosci. Biotechnol. Biochem.*, **56**, 842–844.